

CORRECTED
VERSION*

CORRECTED
VERSION**

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

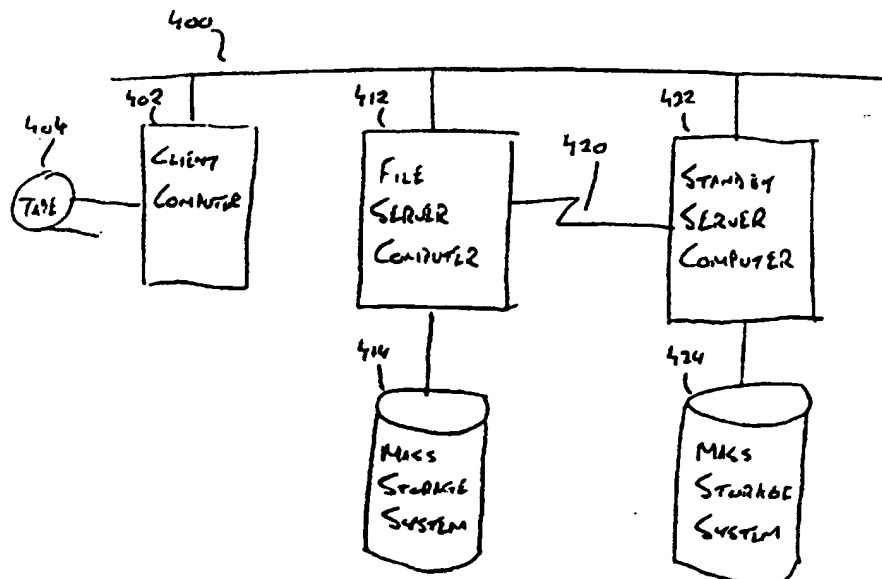


B2

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : G06F 12/16, 12/08		A1	(11) International Publication Number: WO 96/12232
			(43) International Publication Date: 25 April 1996 (25.04.96)
(21) International Application Number: PCT/US95/13324		(81) Designated States: AL, AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TT, UA, UG, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, LS, MW, SD, SZ, UG).	
(22) International Filing Date: 10 October 1995 (10.10.95)		Published With international search report. With amended claims. Date of publication of the amended claims: 29 August 1996 (29.08.96)	
(30) Priority Data: 08/322,697 13 October 1994 (13.10.94) US			
(71) Applicant (for all designated States except US): VINCA CORPORATION [US/US]; 4000 Central Park East, 1815 South State Street, Orem, UT 84058 (US).			
(72) Inventors; and (75) Inventors/Applicants (for US only): OHRAN, Richard, S. [US/US]; 71 West 4750 North, Provo, UT 84604 (US). OHRAN, Michael, R. [US/US]; 109 South 200 East, Orem, UT 84058 (US).			
(74) Agents: CHRISTIANSEN, Jon, C. et al.; Van Cott, Bagley, Cornwall & McCarthy, Suite 1600, 50 South Main Street, P.O. Box 45340, Salt Lake City, UT 84145-0340 (US).			

(54) Title: SNAPSHOT OF DATA STORED ON A MASS STORAGE SYSTEM



(57) Abstract

A method for providing a static snapshot, or image, of data stored on a mass storage system (104). At the start of the method, a preservation memory (106) is cleared and a virtual device is created. Whenever a write is to be performed on the mass storage system (104), a check is made of the preservation memory (106) to determine if it contains a block associated with the mass storage write address. If there is not, a copy of the block in the mass storage system (104) at the block write address is placed in the preservation memory (106). Whenever a read is to be performed on the virtual device, a check is made of the preservation memory (106) to determine if it contains a block associated with the virtual device read address. If there is such a block, that block is returned as the result of the virtual device read. Otherwise, the block at the virtual device block read address is returned as the result.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LJ	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TC	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

AMENDED CLAIMS

[received by the International Bureau on 21 May 1996 (21.05.96);
original claims 1-22 replaced by amended claims 1-22 (11 pages)]

1. A method for providing a static snapshot of data stored on a mass storage system, operating on a computer configuration including:
- a digital computer executing the steps of the method;
 - a mass storage system connected to said digital computer, said mass storage system storing blocks of data having unique addresses; and
 - a preservation memory connected to said digital computer, said preservation memory storing blocks of data associated with said unique addresses;
- the method comprising:
- (A) clearing said preservation memory so that no copies of blocks of data are in said preservation memory;
 - (B) creating a virtual device;
 - (C) whenever a write operation to said mass storage system, said write operation specifying a mass storage write address and data to be written, occurs:
 - (1) if there is not a block of data associated with said mass storage write address in said preservation memory, placing a copy of said block of data located in said mass storage system at said mass storage write address in said preservation memory; and
 - (2) writing said data to be written to said mass storage system at the location specified by said mass storage write address such that a data snapshot is preserved in said preservation memory;
 - and
 - (D) whenever a read operation to said virtual device, said read operation specifying a virtual device read address, occurs:
 - (1) if there is not a block of data associated with said virtual device read address in said preservation

- memory, returning said data of said block of said mass storage system specified by said virtual device read address as the result of said read operation, and
- (2) if there is a block of data associated with said virtual device read address in said preservation memory, returning said block of data from said preservation memory as the result of said read operation, such that a data snapshot is preserved in said preservation memory.
2. A method as in claim 1, wherein said mass storage system comprises one or more disks.
3. A method as in claim 1, wherein said mass storage system is a partition of a disk.
4. A method as in claim 1, wherein said preservation memory is a random-access memory.
5. A method as in claim 1, wherein said preservation memory is one or more disks.
6. A method as in claim 1, wherein said preservation memory is a partition of a disk.
7. A method as in claim 1, wherein said preservation memory is a file stored on mass storage system.
8. A method as in claim 1, the method further comprising:
- (E) whenever a write operation to said virtual device, said write operation specifying a virtual device write address and data to be written, occurs:
- (1) if there is not a block of data associated with said virtual device write address in said preservation memory, placing in said

- preservation memory said data to be written, and
- (2) if there is a block of data associated with said virtual device write address in said preservation memory, replacing in said preservation memory that block of data with said data to be written.
9. A method as in claim 1, the method further comprising:
- (F) whenever a read operation to said mass storage system, said read operation specifying a mass storage read address, occurs, returning said data of said block of said mass storage system specified by said mass storage read address as the result of said read operation.
10. A method as in claim 1, said computer configuration further including a block association memory, said block association memory used to associate blocks stored in said preservation memory with said unique addresses.
11. A method as in claim 10, wherein said block association memory contains entries indicating a unique address in said mass storage system and a location in said preservation memory of a block associated with that unique address.
12. A method as in claim 11, wherein blocks of data in said preservation memory are associated with a unique address by searching said block association memory entries for a matching address.
13. A method as in claim 12, wherein there is not a block of data associated with an address if there is no entry in said block association memory with a matching address.
14. A method as in claim 11, wherein said block association memory contains an entry for each unique

address in said mass storage system indicating a location in said preservation memory of a block associated with that unique address.

15. A method as in claim 12, where a special value for said preservation memory location in said entries indicates that there is not a block of data in said preservation memory associated with that address.

16. A method as in claim 1, wherein said digital computer acts as a file server, and said virtual device is exported to other computers.

17. A method as in claim 16, wherein said mass storage system is exported to other computers.

18. A method as in claim 1, said computer configuration including a second mass storage system, the method further comprising:

creating a second virtual device;
whenever a write operation to said second mass storage system, said write operation specifying a second mass storage write address and data to be written, occurs:

- (1) if there is not a block of data associated with said second mass storage write address in said preservation memory, placing a copy of said block of data located in said second mass storage system at said second mass storage write address in said preservation memory; and
- (2) writing said data to be written to said second mass storage system at the location specified by said second mass storage write address;

and

whenever a read operation to said second virtual device, said read operation specifying a virtual device read address, occurs:

- (1) if there is not a block of data associated with said virtual device read address in said preservation memory, returning said data of said block of said second mass storage system specified by said virtual device read address as the result of said read operation, and
- (2) if there is a block of data associated with said block read address in said preservation memory, returning said block of data from said preservation memory as the result of said read operation.

19. A method as in claim 1, said computer configuration including a second preservation memory, the method further comprising:

- creating a second virtual device;
- whenever a write operation to said mass storage system, said write operation specifying a mass storage write address and data to be written, occurs:
- (1) if there is not a block of data associated with said mass storage write address in said second preservation memory, placing a copy of said block of data located in said mass storage system at said mass storage write address in said second preservation memory; and
 - (2) writing said data to be written to said mass storage system at the location specified by said mass storage write address;

and

whenever a read operation to said second virtual device, said read operation specifying a second virtual device read address, occurs:

- (1) if there is not a block of data associated with said second virtual device read address in said second

- preservation memory, returning said data of said block of said mass storage system specified by said second virtual device read address as the result of said read operation, and
- (2) if there is a block of data associated with said second virtual device read address in said second preservation memory, returning said block of data from said second preservation memory as the result of said read operation.

20. A method for providing a static snapshot of data stored on a mass storage system, the method operating on a computer configuration that includes:

- a digital computer;
- a mass storage system connected to said digital computer, said mass storage system being capable of storing blocks of data having unique addresses; and
- a preservation memory connected to said digital computer, said preservation memory being capable of storing blocks of data associated with said unique addresses;

the method comprising the following steps:

- (A) clearing said preservation memory so that no copies of blocks of data are in said preservation memory;
- (B) creating a virtual device;
- (C) whenever a write operation to said mass storage system occurs that specifies a mass storage write address and data to be written occurs, performing the following:
 - (1) if there is not a block of data associated with said mass storage write address in said preservation memory, placing a copy of said block of data located in said mass storage system at said mass storage

write address in said preservation memory; and

- (2) writing said data to be written to said mass storage system at the location specified by said mass storage write address, such that a data snapshot is preserved in said preservation memory;

and

- (D) whenever a read operation to said virtual device that specifies a virtual device read address occurs, performing the following:

- (1) if there is not a block of data associated with said virtual device read address in said preservation memory, returning said data of said block of said mass storage system specified by said virtual device read address as the result of said read operation, and
- (2) if there is a block of data associated with said virtual device read address in said preservation memory, returning said block of data from said preservation memory as the result of said read operation, such that a data snapshot is preserved in said preservation memory.

21. A method for providing a static snapshot of data stored on a mass storage system, the method operating on a computer configuration that includes:

- a digital computer;
- a mass storage system connected to said digital computer, said mass storage system being capable of storing blocks of data having unique addresses; and
- a preservation memory connected to said digital computer, said preservation memory being capable of storing blocks of data associated with said unique addresses;

the method comprising the following steps:

- (A) clearing said preservation memory so that no copies of blocks of data are in said preservation memory;
- (B) creating a virtual device;
- (C) whenever a write operation to said mass storage system occurs that specifies a mass storage write address and data to be written occurs, performing the following:
 - (1) if there is not a block of data associated with said mass storage write address in said preservation memory, placing a copy of said block of data located in said mass storage system at said mass storage write address in said preservation memory; and
 - (2) writing said data to be written to said mass storage system at the location specified by said mass storage write address, such that a data snapshot is preserved in said preservation memory;
- (D) whenever a read operation to said virtual device that specifies a virtual device read address occurs, performing the following:
 - (1) if there is not a block of data associated with said virtual device read address in said preservation memory, returning said data of said block of said mass storage system specified by said virtual device read address as the result of said read operation, and
 - (2) if there is a block of data associated with said virtual device read address in said preservation memory, returning said block of data from said preservation memory as the result of said read operation, such that a data

- snapshot is preserved in said preservation memory;
- (E) whenever a write operation to said virtual device occurs that specifies a virtual device write address and data to be written, performing the following:
- (1) if there is not a block of data associated with said virtual device write address in said preservation memory, placing in said preservation memory said data to be written, and
 - (2) if there is a block of data associated with said virtual device write address in said preservation memory, replacing in said preservation memory that block of data with said data to be written, such that a data snapshot is preserved in said preservation memory; and
- (F) whenever a read operation to said mass storage system that specifies a mass storage read address occurs, returning said data of said block of said mass storage system specified by said mass storage read address as the result of said read operation, such that a data snapshot is preserved in said preservation memory.

22. A system for providing a static snapshot of data stored on a mass storage system on a computer configuration that includes:

- a digital computer;
- a mass storage system connected to said digital computer, said mass storage system being capable of storing blocks of data having unique addresses; and
- a preservation memory connected to said digital computer, said preservation memory being capable of storing blocks of data associated with said unique addresses;

the system comprising:

- (A) means for clearing said preservation memory so that no copies of blocks of data are in said preservation memory;
- (B) means for creating a virtual device;
- (C) whenever a write operation to said mass storage system occurs that specifies a mass storage write address and data to be written occurs, means for performing the following:
 - (1) if there is not a block of data associated with said mass storage write address in said preservation memory, placing a copy of said block of data located in said mass storage system at said mass storage write address in said preservation memory; and
 - (2) writing said data to be written to said mass storage system at the location specified by said mass storage write address, such that a data snapshot is preserved in said preservation memory;
- (D) whenever a read operation to said virtual device that specifies a virtual device read address occurs, means for performing the following:
 - (1) if there is not a block of data associated with said virtual device read address in said preservation memory, returning said data of said block of said mass storage system specified by said virtual device read address as the result of said read operation, and
 - (2) if there is a block of data associated with said virtual device read address in said preservation memory, returning said block of data from said preservation memory as the result of said read operation, such that a data

- snapshot is preserved in said preservation memory;
- (E) whenever a write operation to said virtual device occurs that specifies a virtual device write address and data to be written, means for performing the following:
- (1) if there is not a block of data associated with said virtual device write address in said preservation memory, placing in said preservation memory said data to be written, such that a data snapshot is preserved in said preservation memory, and
 - (2) if there is a block of data associated with said virtual device write address in said preservation memory, replacing in said preservation memory that block of data with said data to be written, such that a data snapshot is preserved in said preservation memory; and
- (F) whenever a read operation to said mass storage system that specifies a mass storage read address occurs, means for returning said data of said block of said mass storage system specified by said mass storage read address as the result of said read operation, such that a data snapshot is preserved in said preservation memory.

CORRECTED
VERSION*

CORRECTED
VERSION**

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : G06F 12/16, 12/08	A1	(11) International Publication Number: WO 96/12232 (43) International Publication Date: 25 April 1996 (25.04.96)
--	----	---

(21) International Application Number: PCT/US95/13324

(22) International Filing Date: 10 October 1995 (10.10.95)

(30) Priority Data:
08/322,697 13 October 1994 (13.10.94) US

(71) Applicant (for all designated States except US): VINCA CORPORATION [US/US]; 4000 Central Park East, 1815 South State Street, Orem, UT 84058 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): OHRAN, Richard, S. [US/US]; 71 West 4750 North, Provo, UT 84604 (US). OHRAN, Michael, R. [US/US]; 109 South 200 East, Orem, UT 84058 (US).

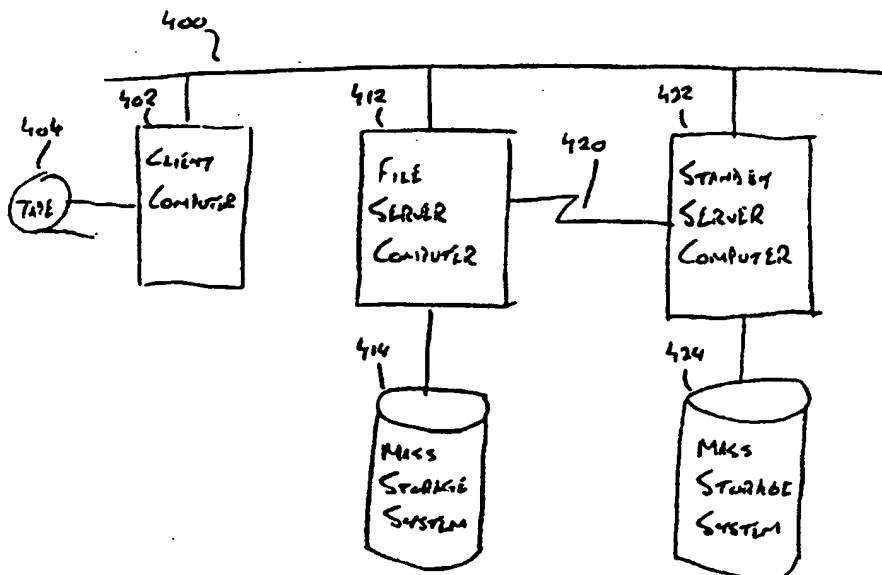
(74) Agents: CHRISTIANSEN, Jon, C. et al.; Van Cott, Bagley, Cornwall & McCarthy, Suite 1600, 50 South Main Street, P.O. Box 45340, Salt Lake City, UT 84145-0340 (US).

(81) Designated States: AL, AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TT, UA, UG, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, LS, MW, SD, SZ, UG).

Published

With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: SNAPSHOT OF DATA STORED ON A MASS STORAGE SYSTEM



(57) Abstract

A method for providing a static snapshot, or image, of data stored on a mass storage system (104). At the start of the method, a preservation memory (106) is cleared and a virtual device is created. Whenever a write is to be performed on the mass storage system (104), a check is made of the preservation memory (106) to determine if it contains a block associated with the mass storage write address. If there is not, a copy of the block in the mass storage system (104) at the block write address is placed in the preservation memory (106). Whenever a read is to be performed on the virtual device, a check is made of the preservation memory (106) to determine if it contains a block associated with the virtual device read address. If there is such a block, that block is returned as the result of the virtual device read. Otherwise, the block at the virtual device block read address is returned as the result.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

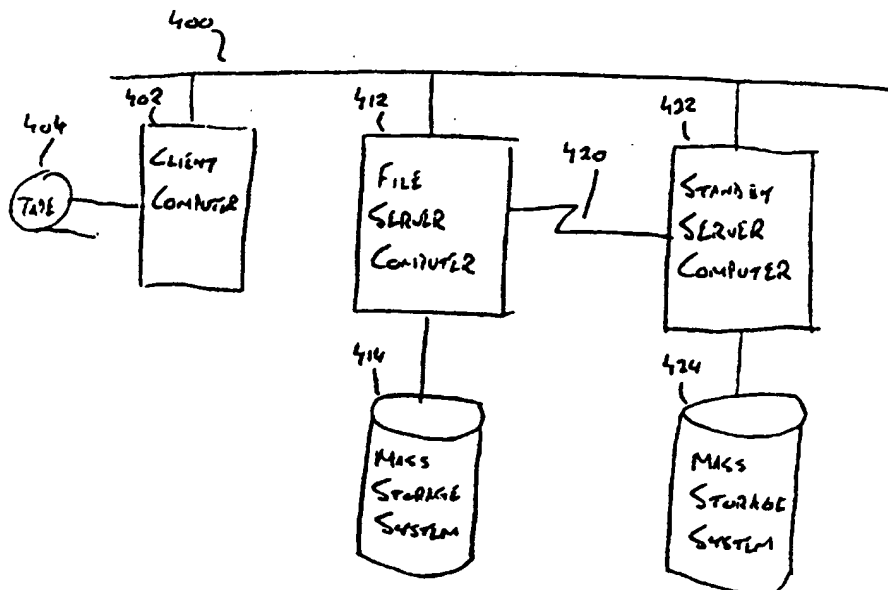
AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LJ	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : G06F 12/16, 12/08		A1	(11) International Publication Number: WO 96/12232
		(43) International Publication Date: 25 April 1996 (25.04.96)	
(21) International Application Number: PCT/US95/13324		(81) Designated States: AL, AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG). Published With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.	
(22) International Filing Date: 10 October 1995 (10.10.95)			
(30) Priority Data: 08/322,697 13 October 1994 (13.10.94) US			
(71) Applicant (for all designated States except US): VINCA CORPORATION [US/US]; 4000 Central Park East, 1815 South State Street, Orem, UT 84058 (US).			
(72) Inventors; and (75) Inventors/Applicants (for US only): OHRAN, Richard, S. [US/US]; 71 West 4750 North, Provo, UT 84604 (US). OHRAN, Michael, R. [US/US]; 109 South 200 East, Orem, UT 84058 (US).			
(74) Agents: CHRISTIANSEN, Jon, C. et al.; Van Cott, Bagley, Cornwall & McCarthy, Suite 1600, 50 South Main Street, P.O. Box 45340, Salt Lake City, UT 84145-0340 (US).			

(54) Title: SNAPSHOT OF DATA STORED ON A MASS STORAGE SYSTEM



(57) Abstract

A method for providing a static snapshot, or image, of data stored on a mass storage system (104). At the start of the method, a preservation memory (106) is cleared and a virtual device is created. Whenever a write is to be performed on the mass storage system (104), a check is made of the preservation memory (106) to determine if it contains a block associated with the mass storage write address. If there is not, a copy of the block in the mass storage system (104) at the block write address is placed in the preservation memory (106). Whenever a read is to be performed on the virtual device, a check is made of the preservation memory (106) to determine if it contains a block associated with the virtual device read address. If there is such a block, that block is returned as the result of the virtual device read. Otherwise, the block at the virtual device block read address is returned as the result.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LJ	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

To all whom it may concern:

Be it known that Richard S. Ohran and Michael R. Ohran, citizens of the United States of America, have invented a new and useful invention entitled METHOD AND SYSTEM FOR PROVIDING A STATIC SNAPSHOT OF DATA STORED ON A MASS STORAGE SYSTEM of which the following comprises a complete specification.

1 However, if the data stored on the mass storage
2 system is being updated by other programs as the backup
3 copy is being made, the image of the data on the mass
4 storage system written to tape may be inconsistent. This
5 is because normal backup techniques either copy the
6 blocks from the mass storage system sequentially to the
7 linear-access tape, or walk the file system stored on the
8 mass storage system, starting with the first block of the
9 first file in the first directory and proceeding in order
10 to the last block of the last file of the last directory.
11 The backup program is not aware of updates performed to a
12 block of the mass storage system after that block has
13 been written to tape.

14 This problem of inconsistent data being written to
15 tape is particularly likely to occur if the mass storage
16 system is being used by a database management system,
17 where an update may involve changing information stored
18 on different parts of the mass storage system. If a
19 database update is made while the backup tape is being
20 written, the image of the database management system
21 written to tape will have the old values for any data
22 already written to tape at the time of the database
23 update, and the new values for any data written to tape
24 following the database update. A restoration based on

1 the tape image of the database would yield an
2 inconsistent database.

3 Horton et al., United States Patent No. 5,089,958,
4 which is hereby incorporated by reference in its entirety
5 for the material disclosed therein, discloses a technique
6 for producing an image of a mass storage system at any
7 point in time after the technique is started. This is
8 done by establishing a base image of the mass storage
9 system at the start of the technique and a log indicating
10 each change made to the mass storage system. An image at
11 any point in time can then be produced by starting with
12 the base image and making all the changes indicated in
13 the log up to that point in time. To improve
14 performance, the Horton system also provides for
15 differential images so that the compilation of changes to
16 form an image does not have to start with the base image.

17 There are two difficulties with using the technique
18 of Horton to provide an image for backup operations.
19 First, the technique is not designed to provide a static
20 snapshot or image of the mass storage system, but to
21 allow an image from any point in time to be created at
22 some later time. This increases the complexity of the
23 technique and requires the compilation of changes
24 whenever a virtual image is desired.

1 The second difficulty with using the technique of
2 Horton is that the log must store a copy of each change
3 made to the mass storage system in order to produce an
4 image of the mass storage system as it was at a specified
5 time. This means that the size of the log can grow
6 without bound, eventually exhausting the space available
7 for its storage. At this point, updates to the mass
8 storage system are not possible without compromising the
9 ability to produce an image from any previous point in
10 time.

11 With many database systems or file systems, certain
12 key blocks (such as master directory blocks) are
13 frequently updated, perhaps with every update to any
14 other block. A copy of these blocks must be written to
15 the log each time they are changed. This will, of
16 course, result in a very large log file, with many of the
17 entries being copies of the key blocks as they changed
18 over time.

19 Another approach to creating a static image of a
20 mass storage system is possible if the mass storage
21 system has the ability to produce a mirror, or identical
22 copy, of one disk's data on a second disk. At the time
23 the static image is needed, mirroring of data is stopped
24 and the mirror disk is used as the static image. When

1 the static image is no longer necessary (for example,
2 when the tape backup has been completed), the two disks
3 are resynchronized, by copying any changes made during
4 the time mirroring was not active to the mirror disk, and
5 mirroring is resumed.

6 This approach also has problems. Unless there are
7 three or more disks mirroring the information on the main
8 disk, when mirroring is stopped to produce the static
9 image there is no longer the redundancy of mirrored disk
10 or disks and updates can be lost if there is a disk
11 failure. Furthermore, it requires an entire disk to be
12 devoted to the storage of the static image.

13 But the major disadvantage of this mirror disk
14 approach is the time necessary to restart mirroring after
15 the static image is no longer needed. This requires
16 updating the mirror disk with all the changes that have
17 been made since mirroring was stopped. If a log of these
18 changes is not available, this means that all the data on
19 the mirror disk must be copied from the disk which has
20 been updated. For large disks such as would be found on
21 a database system, this could take many hours.

22 For more general background reference materials, the
23 reader is directed to United States Patent Nos.
24 3,444,528, 3,533,082, 4,141,066, 4,156,901, 4,164,017,

1 4,191,996, 4,351,023, 4,378,588, 4,403,303, 4,453,215,
2 4,459,658, 4,479,214, 4,483,001, 4,484,275, 4,486,826,
3 4,498,145, 4,507,751, 4,521,847, 4,581,701, 4,607,365,
4 4,608,688, 4,639,856, 4,648,031, 4,652,940, 4,654,819,
5 4,654,857, 4,674,038, 4,703,421, 4,703,481, 4,713,811,
6 4,727,516, 4,736,339, 4,750,177, 4,754,397, 4,878,167,
7 5,307,481, 5,235,700, 5,079,740, 4,823,256, 5,295,258,
8 5,157,663, 4,471,429, 4,530,052, 4,615,001, 4,941,087,
9 4,959,768, 4,979,108, 4,800,488, 4,656,596, 4,866,707,
10 3,771,137, 4,402,046, 4,423,414, 4,430,699, 4,477,882,
11 4,480,304, 4,562,535, 4,604,690, 4,630,224, 4,644,470,
12 4,648,061, 3,754,211, 3,824,547, 4,439,859, 4,445,214,
13 4,691,314, 3,754,211, 4,332,027, 4,516,121, 4,583,089,
14 4,646,300, 5,060,185, 5,123,099, 3,602,900, 3,665,173,
15 3,681,578, 3,735,356, 3,760,364, 3,761,884, 3,810,119,
16 3,828,321, 3,054,560, 3,303,474, 3,544,477, 3,623,014,
17 3,636,331, 3,725,861, 3,803,568, 4,012,717, 4,076,961,
18 4,099,241, 4,118,772, 4,358,823, 4,359,718, 4,455,645,
19 4,477,895, 4,589,090, 4,590,554, 4,610,013, 4,623,883,
20 4,639,852, 4,654,846, 4,680,581, 3,557,315, 3,636,331,
21 3,810,121, 3,812,469, 3,820,085, 3,828,321, 3,864,670,
22 3,865,999, 3,889,237, 3,959,638, 3,991,407, 4,073,005,
23 4,099,241, 4,152,764, 4,208,715, 4,228,503, 4,257,009,
24 4,270,167, 4,282,572, 4,318,173, 4,358,823, 4,371,754,

1 4,403,286, 4,428,044, 4,455,601, 4,530,051, 4,590,554,
2 and 4,628,508 each of which is hereby incorporated by
3 reference in their entirety for the material disclosed
4 therein. The reader's attention is also directed to the
5 following publications: Lyon, "Tandem's Remote Data
6 Facility," IEEE (1990); and Molina et al., "Issues in
7 Disaster Recovery," IEEE (1990), each of which is
8 incorporated by reference in its entirety for the
9 material disclosed therein.

11 Summary of the Invention

12 It is an object of this invention to provide a
13 static image of data stored on a mass storage system as
14 it existed at a particular point in time.

15 This is accomplished by creating a virtual device
16 that will appear as a mass storage device containing the
17 static image. Write operations to the mass storage
18 system are also intercepted by the method. Copies of
19 blocks on the mass storage system are placed in a
20 preservation memory whenever they are going to be changed
21 by a write operation, unless an entry for that block is
22 already in the preservation memory. During a read of the
23 virtual device, the preservation memory is first checked,
24 either directly or using a table of contents of the

1 preservation memory, to see if it contains a copy of the
2 block from the specified location. If the preservation
3 memory has such a copy, that copy is returned as the
4 result of the read. Otherwise, the block is read from the
5 mass storage system.

6 It is a further object of the invention to reduce
7 the amount of storage required to provide the static
8 image. The technique of Horton requires the storage of
9 all changes from the time the technique is started. The
10 mirror disk technique requires storage equal to the size
11 of the mass storage being imaged. In contrast, the
12 method of the invention only requires storage equal to
13 the number of mass storage blocks that have been changed
14 since the static image was created.

15 It is a further object of the invention to reduce
16 the time necessary for generating the static image and
17 for returning to normal operation when the static image
18 is no longer needed. Unlike the technique of Horton,
19 where the static image at a particular time needs to be
20 compiled from the base image and log entries, all that is
21 necessary for creating a static image using the method of
22 this invention is to create the virtual device and
23 establish the interception of writes to the mass storage
24 system. No copying of data or compilation of an image is

1 necessary.

2 When the static image is no longer necessary, the
3 virtual device is removed from the system and the
4 contents of the preservation memory deleted if recovery
5 of that space is desirable. No synchronization to the
6 mass storage system is necessary nor is it necessary to
7 merge updates into a base image to create a new base
8 image.

9 These and other features of the invention will be
10 more readily understood upon consideration of the
11 attached drawings and of the following detailed
12 description of those drawings and the presently preferred
13 embodiments of the invention.

14
15
16 Brief Description of the Drawings

17 Figure 1 is a block diagram of a representative
18 digital computer configuration on which the preferred
19 embodiment of the invention operates.

20 Figure 2 is a flow diagram showing the preferred
21 steps of the method of the invention.

22 Figure 3 is a block diagram of a client-server
23 configuration using the preferred method.

24 Figure 4 is a block diagram of the currently-

1 preferred embodiment of the method in a client-server
2 configuration.

3 Figure 5 is variant of the configuration of Figure
4 4.

5
6 **Detailed Description of the Invention**

7 Referring to Figure 1, which illustrates a
8 representative computer configuration on which the method
9 of the invention runs, it can be seen that digital
10 computer 102 has connected to it mass storage system 104
11 and preservation memory 106. In some embodiments of the
12 invention, association memory 108 may also be connected
13 to digital computer 102.

14 Mass storage system 104 can be any writable block-
15 addressable storage system, such as one or more disks or
16 a partition of a disk. (If mass storage system 104 were
17 not writable, such as a CD-ROM, its contents would not
18 change and there would be no need for the invention of
19 this application.) A partition of a disk can be a fixed
20 area of a disk. The disks can store their information
21 using magnetic, optical, or any other technique that
22 allows writing and reading of data without departing from
23 the scope and spirit of this invention.

24 In the currently-preferred embodiment of the

1 invention, digital computer 102 is a PC-compatible
2 computer based on an Intel X86 series or compatible
3 processor and mass storage device 104 is a SCSI or IDE
4 magnetic disk connected to digital computer 102 through
5 an appropriate controller.

6 Preservation memory 106 can be an area in the
7 random-access memory (RAM) of digital computer 102, one
8 or more disks, a partition of a disk, or a file stored on
9 a disk. Optimal selection of the implementation of
10 preservation memory 106 depends of the number of blocks
11 of the mass storage system that will be changed during
12 the time the static image is needed. The use of RAM
13 provides faster performance, but may limit the number of
14 entries in the preservation memory. If the preservation
15 memory runs out of space when a new entry must be stored,
16 the method fails and the static image is no longer
17 available or remains in the state it was at the time the
18 preservation memory ran out of space. It is important to
19 note that if this occurs, no data from the mass storage
20 system is lost, and the method can be restarted to
21 produce a new static image.

22 Referring to Figure 2, which is a flow diagram
23 showing the steps of the method, the method starts at
24 step 202 when a static image of the mass storage system

1 is desired. This can be indicated by the running of a
2 special program, an operating system call, or an operator
3 command, as appropriate for the operating system and
4 application.

5 In step 202, preservation memory 106 is cleared. In
6 general, this will consist of setting the control
7 information describing the contents of preservation
8 memory 106 to indicate that there are no valid entries in
9 preservation memory 106.

10 In step 204, a virtual device appearing as a mass
11 storage device is created. The method for creating a
12 virtual device will depend on the particular operating
13 system running on digital computer 102, but will be known
14 by one skilled in the art of that particular operating
15 system. In addition, it may be necessary in step 204 to
16 configure the operating system so that the method of this
17 invention intercepts any read or write operation directed
18 to mass storage system 104. Again, how this is done will
19 be dependent on the particular operating system running
20 on digital computer 102.

21 In step 206, the method waits until there is a write
22 operation directed to mass storage system 104 or a read
23 operation directed to the virtual device created in step
24 204. In variants of the method, step 206 also reacts to

1 a read operation directed to mass storage system 104 or a
2 write operation directed to the virtual device created in
3 step 204.

4 If the operation is a write to mass storage system
5 104, step 210 is entered. Using the mass storage write
6 address specified in the write operation, step 210
7 determines if there is a block of data associated with
8 that mass storage write address in preservation memory
9 106. If there isn't, step 212 of the method is executed.
10 Otherwise, step 212 is skipped and step 214 is executed.

11 There are a number of ways for determining whether
12 there is a block of data associated with the mass storage
13 address in the preservation memory 106. In the
14 currently-preferred embodiment of the invention, there is
15 a block association memory 108 also connected to digital
16 computer 102. (Block association memory 108 may be a
17 separate memory connected to digital computer 102, or may
18 be a portion of the RAM of digital computer 102.) Block
19 association memory 108 is used to associate blocks stored
20 in preservation memory 106 with the unique addresses of
21 blocks on mass storage system 104. Block association
22 memory 108 does this by containing entries indicating a
23 unique address and the location in preservation memory
24 106 for the block associated with that unique address.

1 Entries in block association memory 108 can be
2 stored unordered, in which case they must be linearly
3 searched for a matching unique address. If no entry is
4 found with a matching address, there is not a block in
5 preservation memory 106 associated with that address.
6 Alternatively, the entries could be stored ordered by
7 unique addresses, in which case a binary search could be
8 used to locate the matching entry in block association
9 memory 108. A hashing scheme could also be used to find
10 a matching entry.

11 The block association memory 108 can also be
12 organized as an array with an element for each unique
13 address of mass storage system 104. Each element of the
14 array stores a preservation memory location, or a special
15 value that indicates that there is not a block in
16 preservation memory 106 associated with that unique
17 address.

18 The selection of a technique for storing entries in
19 block association memory 108 depends on the
20 characteristics of accessing the entries. Using an array
21 provides the highest speed for accessing an entry or
22 adding an entry corresponding to a block just copied into
23 preservation memory 106, at the expense of a large block
24 association memory 108. Ordering the entries by unique

1 address provides faster access than for unordered
2 entries, but requires more time when an entry is added to
3 block association memory 108. In the currently-preferred
4 embodiment, entries are stored unordered in block
5 association memory 108.

6 In this discussion, the term block refers to the
7 data stored at a particular location in mass storage
8 system 104 or preservation memory 106. Blocks are
9 generally of a fixed size (e.g. 512 bytes for disks used
10 with MS-DOS), although blocks of different sizes, or
11 variable sizes, are within the scope of this invention.
12 On mass storage system 104, each block has a unique
13 address, specified in read or write operations. A block
14 in preservation memory 106 is a copy of a block of data
15 stored in mass storage system 104, and that block in
16 preservation memory 106 is associated with the unique
17 address of the block in mass storage system 104 of which
18 it is a copy.

19 For efficiency, it may be convenient to treat one or
20 more contiguous blocks on mass storage system 104 as if
21 it were a single, large block. Often operating systems
22 perform their mass storage operations on contiguous
23 blocks (called clusters in MS-DOS). The extensions to
24 handle clusters of blocks should be clear to one with

1 ordinary skills in computer programming.

2 If block association memory 108 is being used, step
3 202 (clearing preservation memory 106) consists of
4 removing all entries from block association memory 108 or
5 setting them to the special entry that indicates that
6 there is no block in preservation memory 106 associated
7 with each unique address.

8 Returning to Figure 2, step 212 is executed if there
9 is not a block associated with the mass storage write
10 address in preservation memory 106. Step 212 places a
11 copy of the block of data currently located at the mass
12 storage write address in preservation memory 106,
13 updating block association memory 108 as necessary. It
14 is important to note that step 212 will be executed at
15 most once for each unique address on mass storage system
16 104, since the next time step 210 tests to see if there
17 is a block in preservation memory 106 associated with
18 that mass storage write address it will find the copy
19 made by step 212. Because of this, preservation memory
20 106 will contain only copies of blocks as they were when
21 the method was started.

22 In step 214, the data to be written by the mass
23 storage write operation is written to the location on
24 mass storage system 104 specified by the mass storage

1 write address. This completes the steps for a mass
2 storage write, and step 206 is reentered to wait for the
3 next operation.

4 If the operation is a virtual device read, step 220
5 is entered. Again, a check is made to determine if a
6 block associated with the virtual device read address is
7 in preservation memory 106. If there is such a block,
8 step 224 is executed. If not, step 222 is executed.

9 Step 222 returns the data from the block in mass
10 storage system 104 specified by the virtual device read
11 address as the result of the read operation. Step 224
12 returns the block from preservation memory 106 associated
13 with the virtual address read address as the result of
14 the read operation. This completes the steps for a
15 virtual device read, and step 206 is reentered to wait
16 for the next operation.

17 If the operation is a mass storage read, step 230 is
18 entered, which returns the data from the block of mass
19 storage system 104 specified by the mass storage read
20 address as the result of the read operation. This
21 completes the steps for a mass storage read, and step 206
22 is reentered to wait for the next operation.

23 It may be desirable to allow write operations to the
24 virtual device, changing the image as specified by the

1 write operations. For example, it may be necessary to
2 write a different label or other control information on
3 the virtual device image so the operating system can
4 differentiate it from mass storage system 104.

5 If the operation is a virtual device write, step 240
6 is entered. Step 240 checks to see if the virtual device
7 is read-only, and if it is step 242 is entered to return
8 an appropriate error indication to the operating system
9 or user.

10 Step 244 checks to determine if a block associated
11 with the virtual device write address is in preservation
12 memory 106. If there is such a block, step 248 is
13 executed. If not, step 246 is executed. In step 246,
14 the data from the virtual device write operation is
15 placed in preservation memory 106, associated with the
16 virtual device write address from the virtual device
17 write operation. Block association memory 108 is updated
18 as necessary. In step 248, the data from the virtual
19 device write operation replaces the block associated with
20 the virtual device write address of the virtual device
21 write operation. This completes the steps of the virtual
22 device write, and step 206 is reentered to wait for the
23 next operation.

24 While the description above describes the basic

1 operation of the method of the invention, there are a
2 number of other embodiments possible. For example, the
3 same preservation memory 106 can be shared so that a
4 second virtual device provides a snapshot image of a
5 second mass storage system. In another embodiment, a
6 second preservation memory and second virtual device can
7 be used to provide a second image whose snapshot was
8 taken at a different time of mass storage system 104.

9 The computer system running the method of the
10 invention can also be used as a file server for client
11 computers connected to it by a network or other means.
12 As a file server, it can export its mass storage system,
13 the virtual device created by the method, or both. Such
14 as system is illustrated in Figure 3.

15 File server computer 312, with mass storage system
16 314, runs the method of the invention. It exports the
17 virtual device (and probably mass storage system 314) to
18 client computer 302, communicating over network 300.
19 Computer 302 can run a tape backup program that copies
20 the information from the exported virtual device to tape
21 drive 304. No change is necessary for the tape backup
22 program running on client computer 302, which sees the
23 virtual device as just another mass storage device.

24 Figure 4 illustrates the currently-preferred

1 configuration for running the method of the invention.
2 Network 400 connects client computer 402, with tape drive
3 404, to file server computer 412, with mass storage
4 system 414. File server computer communicates with
5 standby server computer 422 over data link 420. Standby
6 server computer 422 has mass storage system 424. Through
7 software running on file server computer 412 and standby
8 server computer 422, as described in United States Patent
9 application serial number 08/094,744, filed on July 20,
10 1993 and entitled "METHOD FOR RAPID RECOVERY FROM A
11 NETWORK FILE SERVER FAILURE" (which is hereby
12 incorporated by reference in its entirety), mass storage
13 system 424 appears as a disk to file server computer 412
14 and mirrors the data on mass storage system 414. In the
15 event of a failure of either file server computer 412 or
16 mass storage system 414, standby computer 422 can be
17 restarted as the file server.

18 In the configuration of Figure 4, standby server 422
19 runs the method of the invention, and can export the
20 virtual device either to file server computer 412, which
21 can then export it to client computers on network 400, or
22 standby server 422 can directly export the virtual device
23 to client computers. The virtual device can also be
24 accessed by programs running on standby server 422.

1 Figure 5 illustrates a variant of the configuration
2 of Figure 4. Instead of client computer 502 having a
3 tape drive, as was the case for client computer 402,
4 backup computer 532 has tape drive 534. Backup computer
5 532 communicates with standby server computer 522 over
6 data link 530. Standby server computer exports mass
7 storage system 524 to file server computer 512 (whether
8 mirrored or not). Standby server computer exports the
9 virtual device with the snapshot image of mass storage
10 system 524 to backup computer 532.

11 Backup computer 532 can now copy the snapshot image
12 of mass storage system 524 by reading the virtual device
13 exported to it by standby server computer 522. Neither
14 file server computer 512 nor standby server computer 522
15 has the overhead of the tape backup process, which can
16 result in a degradation of performance if data
17 compression needs to be performed before the data is
18 written to tape. Also, a fault in the tape backup
19 program will not affect either file server computer 512
20 or standby server computer 522.

21 It is to be understood that the above described
22 embodiments are merely illustrative of numerous and
23 varied other embodiments which may constitute
24 applications of the principles of the invention. Such

1 other embodiments may be readily devised by those skilled
2 in the art without departing from the spirit or scope of
3 this invention and it is our intent they be deemed within
4 the scope of our invention.

5
6
7 Claims

8 We claim:

9 1. A method for providing a static snapshot of data
10 stored on a mass storage system, operating on a computer
11 configuration including:

12 a digital computer executing the steps of the
13 method;

14 a mass storage system connected to said digital
15 computer, said mass storage system storing
16 blocks of data having unique addresses; and
17 a preservation memory connected to said digital
18 computer, said preservation memory storing
19 blocks of data associated with said unique
20 addresses;

21 the method comprising:

22 (A) clearing said preservation memory so that no
23 copies of blocks of data are in said
24 preservation memory;

1 (B) creating a virtual device;

2 (C) whenever a write operation to said mass storage
3 system, said write operation specifying a mass
4 storage write address and data to be written,
5 occurs:

6 (1) if there is not a block of data associated
7 with said mass storage write address in
8 said preservation memory, placing a copy
9 of said block of data located in said mass
10 storage system at said mass storage write
11 address in said preservation memory; and

12 (2) writing said data to be written to said
13 mass storage system at the location
14 specified by said mass storage write
15 address;

16 and

17 (D) whenever a read operation to said virtual
18 device, said read operation specifying a
19 virtual device read address, occurs:

20 (1) if there is not a block of data associated
21 with said virtual device read address in
22 said preservation memory, returning said
23 data of said block of said mass storage
24 system specified by said virtual device

1 read address as the result of said read
2 operation, and

3 (2) if there is a block of data associated
4 with said virtual device read address in
5 said preservation memory, returning said
6 block of data from said preservation
7 memory as the result of said read
8 operation.

9
10 2. A method as in claim 1, wherein said mass storage
11 system comprises one or more disks.

12
13 3. A method as in claim 1, wherein said mass storage
14 system is a partition of a disk.

15
16 4. A method as in claim 1, wherein said preservation
17 memory is a random-access memory.

18
19 5. A method as in claim 1, wherein said preservation
20 memory is one or more disks.

21
22 6. A method as in claim 1, wherein said preservation
23 memory is a partition of a disk.

1 7. A method as in claim 1, wherein said preservation
2 memory is a file stored on mass storage system.
3

4 8. A method as in claim 1, the method further
5 comprising:

6 (E) whenever a write operation to said virtual
7 device, said write operation specifying a
8 virtual device write address and data to be
9 written, occurs:

10 (1) if there is not a block of data associated
11 with said virtual device write address in
12 said preservation memory, placing in said
13 preservation memory said data to be
14 written, and

15 (2) if there is a block of data associated
16 with said virtual device write address in
17 said preservation memory, replacing in
18 said preservation memory that block of
19 data with said data to be written.
20

21 9. A method as in claim 1, the method further
22 comprising:

23 (F) whenever a read operation to said mass storage
24 system, said read operation specifying a mass

1 storage read address, occurs, returning said
2 data of said block of said mass storage system
3 specified by said mass storage read address as
4 the result of said read operation.
5

6 10. A method as in claim 1, said computer configuration
7 further including a block association memory, said block
8 association memory used to associate blocks stored in
9 said preservation memory with said unique addresses.
10

11 11. A method as in claim 10, wherein said block
12 association memory contains entries indicating a unique
13 address in said mass storage system and a location in
14 said preservation memory of a block associated with that
15 unique address.
16

17 12. A method as in claim 11, wherein blocks of data in
18 said preservation memory are associated with a unique
19 address by searching said block association memory
20 entries for a matching address.
21

22 13. A method as in claim 12, wherein there is not a
23 block of data associated with an address if there is no
24 entry in said block association memory with a matching address.

1 14. A method as in claim 11, wherein said block
2 association memory contains an entry for each unique
3 address in said mass storage system indicating a location
4 in said preservation memory of a block associated with
5 that unique address.

6
7 15. A method as in claim 12, where a special value for
8 said preservation memory location in said entries
9 indicates that there is not a block of data in said
10 preservation memory associated with that address.

11
12 16. A method as in claim 1, wherein said digital
13 computer acts as a file server, and said virtual device
14 is exported to other computers.

15
16 17. A method as in claim 16, wherein said mass storage
17 system is exported to other computers.

18
19 18. A method as in claim 1, said computer configuration
20 including a second mass storage system, the method
21 further comprising:

22 creating a second virtual device;

23 whenever a write operation to said second mass
24 storage system, said write operation specifying

1 a second mass storage write address and data to
2 be written, occurs:

3 (1) if there is not a block of data associated
4 with said second mass storage write
5 address in said preservation memory,
6 placing a copy of said block of data
7 located in said second mass storage system
8 at said second mass storage write address
9 in said preservation memory; and

10 (2) writing said data to be written to said
11 second mass storage system at the location
12 specified by said second mass storage
13 write address;

14 and

15 whenever a read operation to said second virtual
16 device, said read operation specifying a
17 virtual device read address, occurs:

18 (1) if there is not a block of data associated
19 with said virtual device read address in
20 said preservation memory, returning said
21 data of said block of said second mass
22 storage system specified by said virtual
23 device read address as the result of said
24 read operation, and

- 1 (2) if there is a block of data associated
2 with said block read address in said
3 preservation memory, returning said block
4 of data from said preservation memory as
5 the result of said read operation.
6

7 19. A method as in claim 1, said computer configuration
8 including a second preservation memory, the method
9 further comprising:

10 creating a second virtual device;
11 whenever a write operation to said mass storage
12 system, said write operation specifying a mass
13 storage write address and data to be written,
14 occurs:

- 15 (1) if there is not a block of data associated
16 with said mass storage write address in
17 said second preservation memory, placing a
18 copy of said block of data located in said
19 mass storage system at said mass storage
20 write address in said second preservation
21 memory; and
22 (2) writing said data to be written to said
23 mass storage system at the location
24 specified by said mass storage write

1 address;
2 and
3 whenever a read operation to said second virtual
4 device, said read operation specifying a second
5 virtual device read address, occurs:
6 (1) if there is not a block of data associated
7 with said second virtual device read
8 address in said second preservation
9 memory, returning said data of said block
10 of said mass storage system specified by
11 said second virtual device read address as
12 the result of said read operation, and
13 (2) if there is a block of data associated
14 with said second virtual device read
15 address in said second preservation
16 memory, returning said block of data from
17 said second preservation memory as the
18 result of said read operation.

19
20 20. A method for providing a static snapshot of data
21 stored on a mass storage system, the method operating on
22 a computer configuration that includes:
23 a digital computer;
24 a mass storage system connected to said digital

1 computer, said mass storage system being
2 capable of storing blocks of data having unique
3 addresses; and

4 a preservation memory connected to said digital
5 computer, said preservation memory being
6 capable of storing blocks of data associated
7 with said unique addresses;

8 the method comprising the following steps:

9 (A) clearing said preservation memory so that no
10 copies of blocks of data are in said
11 preservation memory;

12 (B) creating a virtual device;

13 (C) whenever a write operation to said mass storage
14 system occurs that specifies a mass storage
15 write address and data to be written occurs,
16 performing the following:

- 17 (1) if there is not a block of data associated
18 with said mass storage write address in
19 said preservation memory, placing a copy
20 of said block of data located in said mass
21 storage system at said mass storage write
22 address in said preservation memory; and
23 (2) writing said data to be written to said
24 mass storage system at the location

1 specified by said mass storage write
2 address;
3 and
4 (D) whenever a read operation to said virtual
5 device that specifies a virtual device read
6 address occurs, performing the following:
7 (1) if there is not a block of data associated
8 with said virtual device read address in
9 said preservation memory, returning said
10 data of said block of said mass storage
11 system specified by said virtual device
12 read address as the result of said read
13 operation, and
14 (2) if there is a block of data associated
15 with said virtual device read address in
16 said preservation memory, returning said
17 block of data from said preservation
18 memory as the result of said read
19 operation.
20
21 21. A method for providing a static snapshot of data
22 stored on a mass storage system, the method operating on
23 a computer configuration that includes:
24 a digital computer;

1 a mass storage system connected to said digital
2 computer, said mass storage system being
3 capable of storing blocks of data having unique
4 addresses; and

5 a preservation memory connected to said digital
6 computer, said preservation memory being
7 capable of storing blocks of data associated
8 with said unique addresses;

9 the method comprising the following steps:

10 (A) clearing said preservation memory so that no
11 copies of blocks of data are in said
12 preservation memory;

13 (B) creating a virtual device;

14 (C) whenever a write operation to said mass storage
15 system occurs that specifies a mass storage
16 write address and data to be written occurs,
17 performing the following:

18 (1) if there is not a block of data associated
19 with said mass storage write address in
20 said preservation memory, placing a copy
21 of said block of data located in said mass
22 storage system at said mass storage write
23 address in said preservation memory; and

24 (2) writing said data to be written to said

1 mass storage system at the location
2 specified by said mass storage write
3 address;

4 (D) whenever a read operation to said virtual
5 device that specifies a virtual device read
6 address occurs, performing the following:

7 (1) if there is not a block of data associated
8 with said virtual device read address in
9 said preservation memory, returning said
10 data of said block of said mass storage
11 system specified by said virtual device
12 read address as the result of said read
13 operation, and

14 (2) if there is a block of data associated
15 with said virtual device read address in
16 said preservation memory, returning said
17 block of data from said preservation
18 memory as the result of said read
19 operation;

20 (E) whenever a write operation to said virtual
21 device occurs that specifies a virtual device
22 write address and data to be written,
23 performing the following:

24 (1) if there is not a block of data associated

1 with said virtual device write address in
2 said preservation memory, placing in said
3 preservation memory said data to be
4 written, and

5 (2) if there is a block of data associated
6 with said virtual device write address in
7 said preservation memory, replacing in
8 said preservation memory that block of
9 data with said data to be written; and

10 (F) whenever a read operation to said mass storage
11 system that specifies a mass storage read
12 address occurs, returning said data of said
13 block of said mass storage system specified by
14 said mass storage read address as the result of
15 said read operation.
16

17 22. A system for providing a static snapshot of data
18 stored on a mass storage system on a computer
19 configuration that includes:

20 a digital computer;

21 a mass storage system connected to said digital
22 computer, said mass storage system being
23 capable of storing blocks of data having unique
24 addresses; and

1 a preservation memory connected to said digital
2 computer, said preservation memory being
3 capable of storing blocks of data associated
4 with said unique addresses;

5 the system comprising:

6 (A) means for clearing said preservation memory so
7 that no copies of blocks of data are in said
8 preservation memory;

9 (B) means for creating a virtual device;

10 (C) whenever a write operation to said mass storage
11 system occurs that specifies a mass storage
12 write address and data to be written occurs,
13 means for performing the following:

14 (1) if there is not a block of data associated
15 with said mass storage write address in
16 said preservation memory, placing a copy
17 of said block of data located in said mass
18 storage system at said mass storage write
19 address in said preservation memory; and

20 (2) writing said data to be written to said
21 mass storage system at the location
22 specified by said mass storage write
23 address;

24 (D) whenever a read operation to said virtual

1 device that specifies a virtual device read
2 address occurs, means for performing the
3 following:

- 4 (1) if there is not a block of data associated
5 with said virtual device read address in
6 said preservation memory, returning said
7 data of said block of said mass storage
8 system specified by said virtual device
9 read address as the result of said read
10 operation, and
11 (2) if there is a block of data associated
12 with said virtual device read address in
13 said preservation memory, returning said
14 block of data from said preservation
15 memory as the result of said read
16 operation;

17 (E) whenever a write operation to said virtual
18 device occurs that specifies a virtual device
19 write address and data to be written, means for
20 performing the following:

- 21 (1) if there is not a block of data associated
22 with said virtual device write address in
23 said preservation memory, placing in said
24 preservation memory said data to be

1 written, and

2 (2) if there is a block of data associated
3 with said virtual device write address in
4 said preservation memory, replacing in
5 said preservation memory that block of
6 data with said data to be written; and

7 (F) whenever a read operation to said mass storage
8 system that specifies a mass storage read
9 address occurs, means for returning said data
10 of said block of said mass storage system
11 specified by said mass storage read address as
12 the result of said read operation.
13

1 / 4

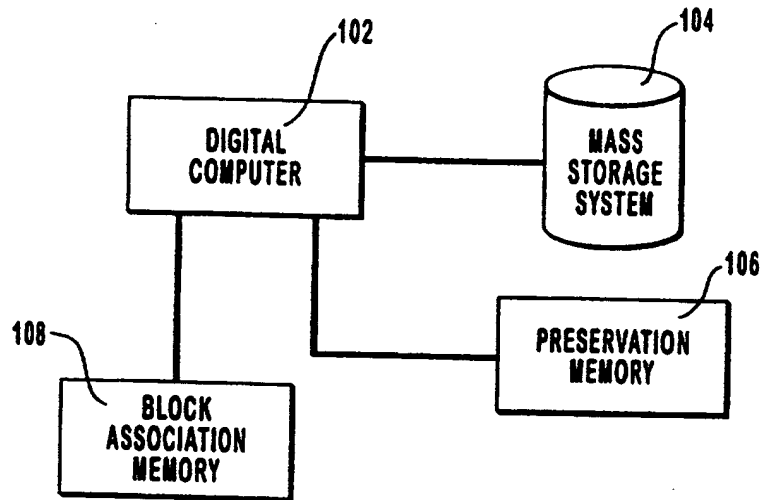


FIG. 1

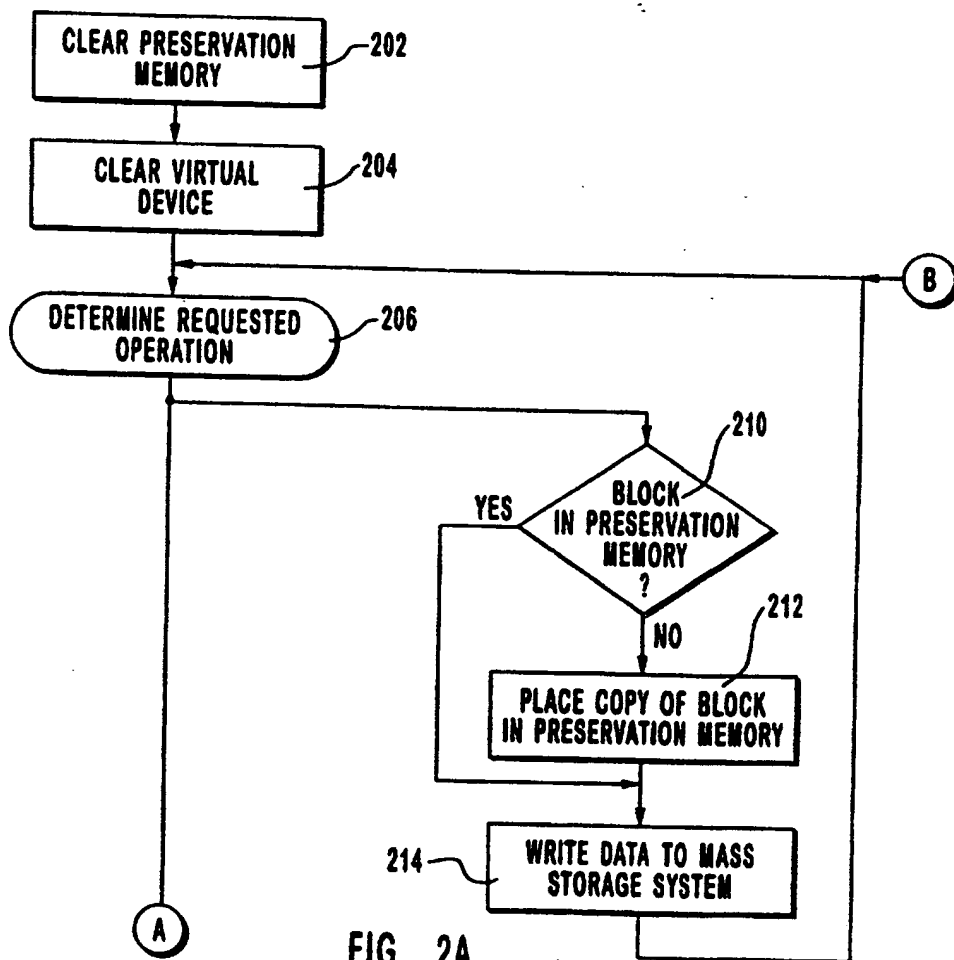


FIG. 2A

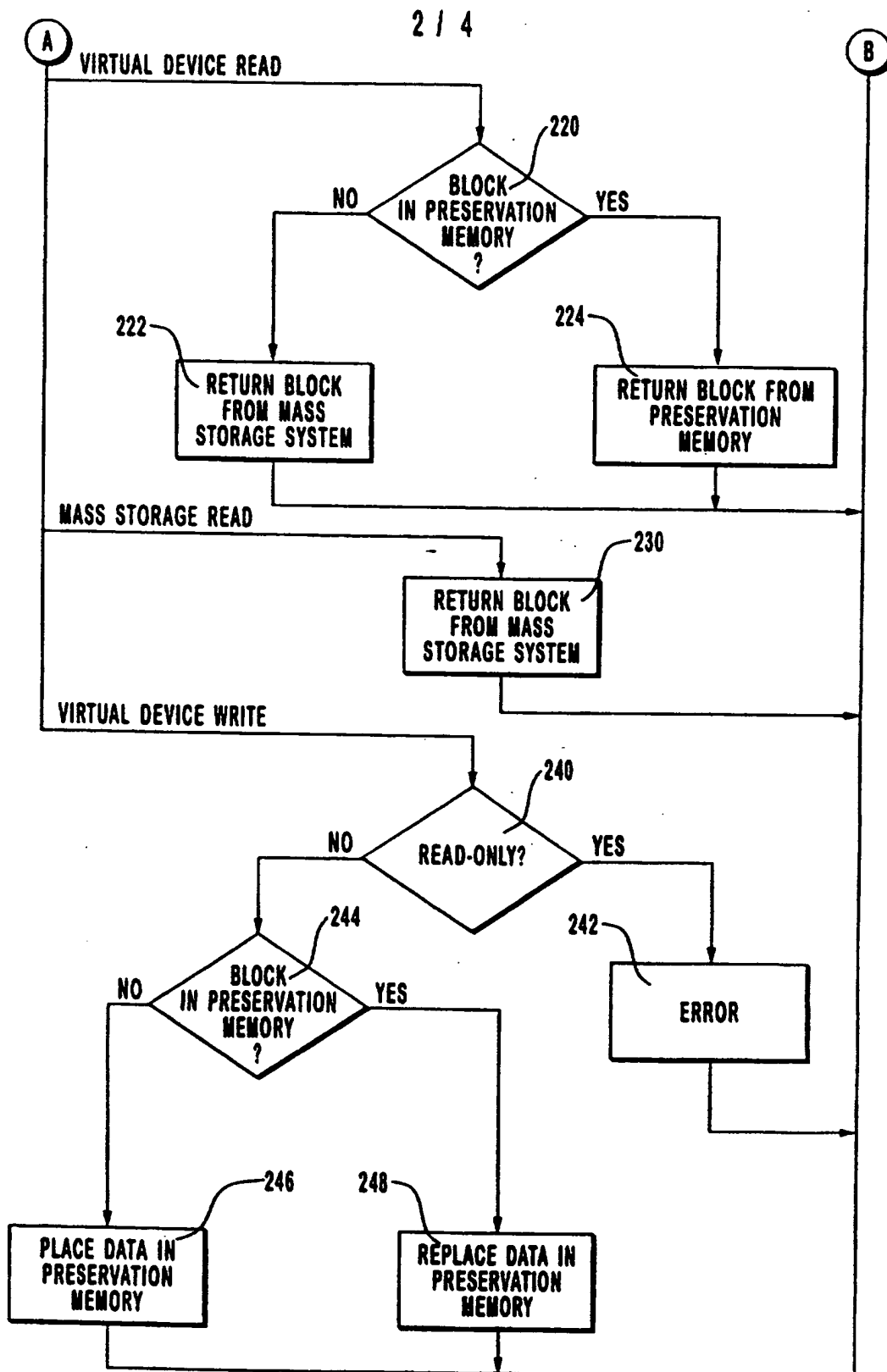


FIG. 2B
SUBSTITUTE SHEET (RULE 26)

3 / 4

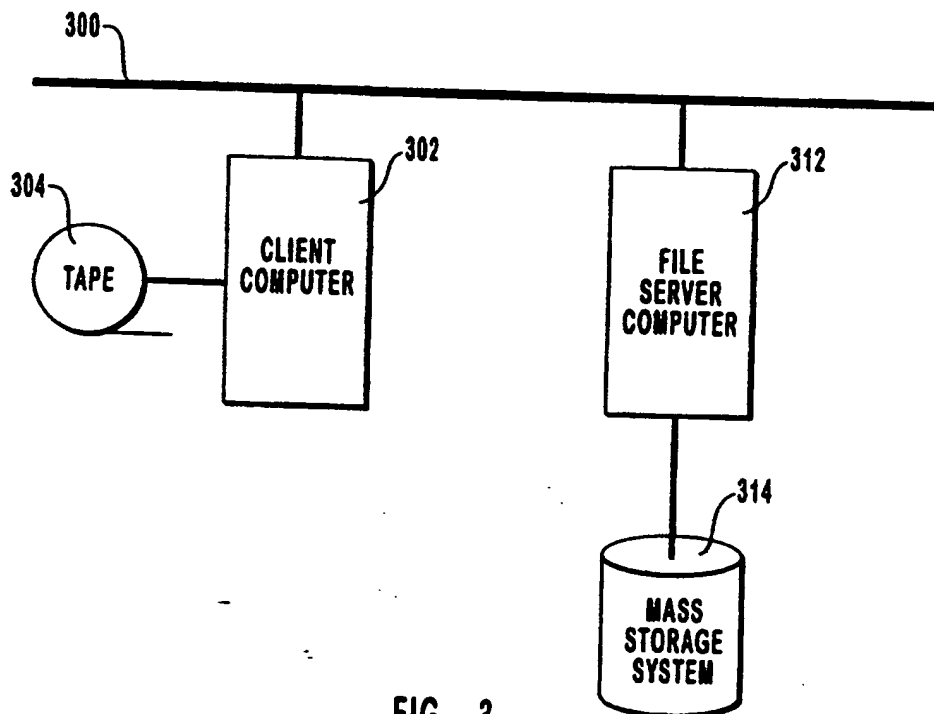


FIG. 3

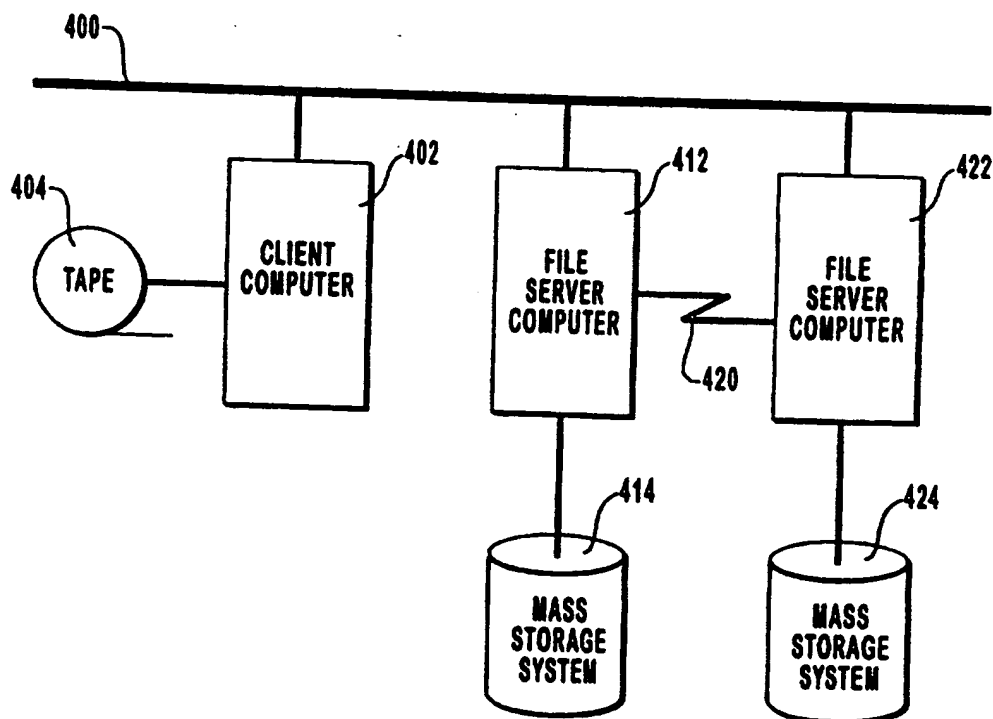


FIG. 4

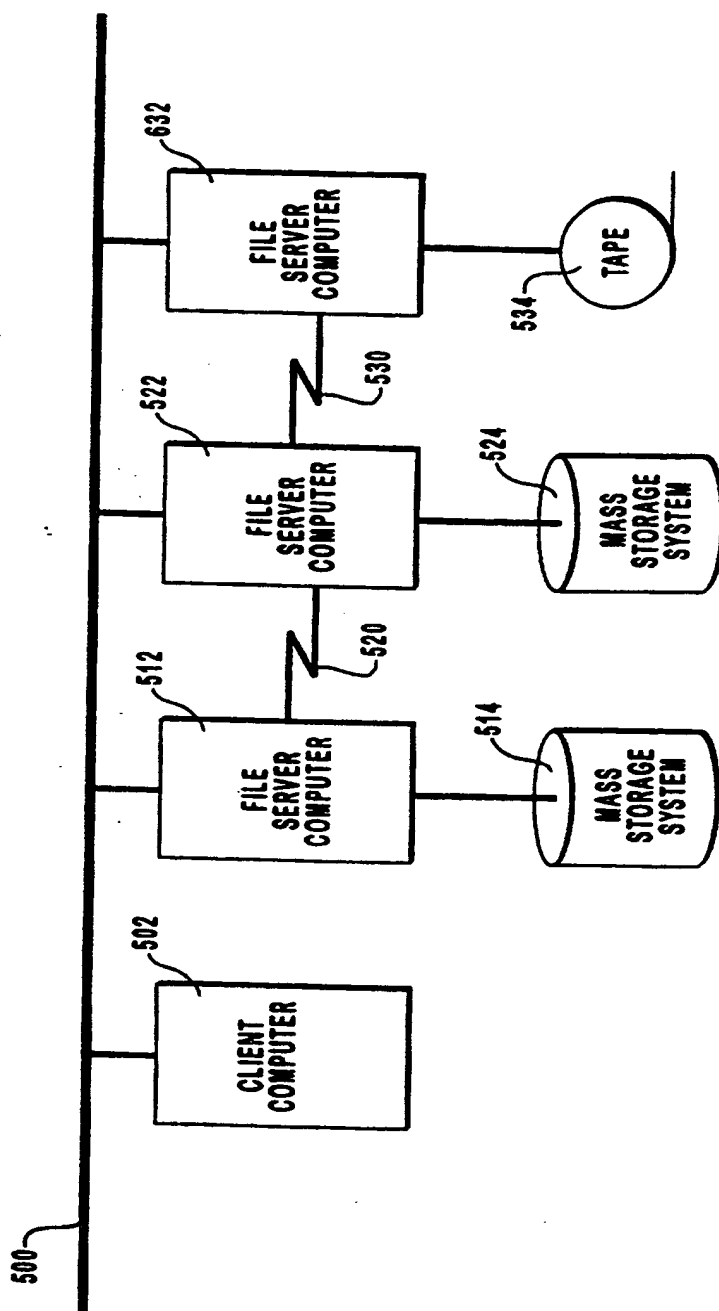


FIG. 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/13324

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : G06F 12/16, 12/08

US CL : 395/468, 488, 462, 412

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 395/468-470, 488-489, 462, 412-413, 417 and 419.

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	US, A, 5,379,391 (BELSAN ET AL) 03 January 1995, column 1, lines 17-22, 43-49, column 11, line 1 to column 13, line 39. Also, see the abstract.	1-14, 18-22
X,P	US, A, 5,426,747 (WEINREB ET AL) 20 June 1995, column 2, lines 50-60, column 6, line 57 to column 7, line 29, claim 1, column 11, lines 30-38, column 12, lines 10-15, column 14, lines 29-37, column 3, lines 17-21.	1-17, 20-22

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	* T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
* A* document defining the general state of the art which is not considered to be part of particular relevance	* X	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
* E earlier document published on or after the international filing date	* Y	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
* L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	* A*	document member of the same patent family
* O document referring to an oral disclosure, use, exhibition or other means		
* T document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

05 JANUARY 1996

Date of mailing of the international search report

22 MAR 1996

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer
ARIO ETIENNE

Telephone No. (703) 308-7562

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/13324

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

USPTO, Automated Patent Search (APS).
Files: USPAT & JPOAPS

Search terms: snapshot, image, intercept, virtual device, disks, mass storage, cache, flush, clear, files.

PCT

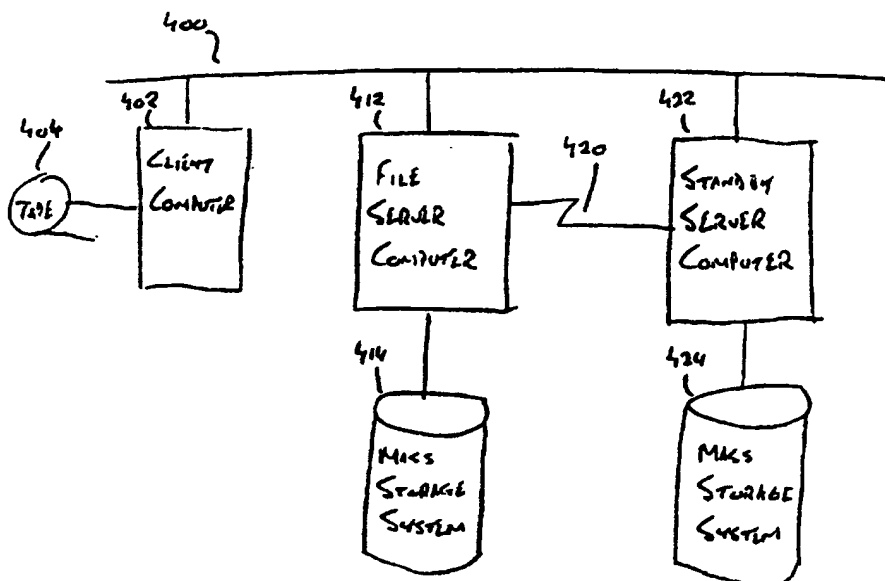
WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : G06F 12/16, 12/08		A1	(11) International Publication Number: WO 96/12232
			(43) International Publication Date: 25 April 1996 (25.04.96)
(21) International Application Number: PCT/US95/13324		(81) Designated States: AL, AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).	
(22) International Filing Date: 10 October 1995 (10.10.95)			
(30) Priority Data: 08/322,697 13 October 1994 (13.10.94) US			
(71) Applicant (for all designated States except US): VINCA CORPORATION [US/US]; 4000 Central Park East, 1815 South State Street, Orem, UT 84058 (US).		Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(72) Inventors; and			
(75) Inventors/Applicants (for US only): OHRAN, Richard, S. [US/US]; 71 West 4750 North, Provo, UT 84604 (US). OHRAN, Michael, R. [US/US]; 109 South 200 East, Orem, UT 84058 (US).			
(74) Agents: CHRISTIANSEN, Jon, C. et al.; Van Cott, Bagley, Cornwall & McCarthy, Suite 1600, 50 South Main Street, P.O. Box 45340, Salt Lake City, UT 84145-0340 (US).			

(54) Title: SNAPSHOT OF DATA STORED ON A MASS STORAGE SYSTEM



(57) Abstract

A method for providing a static snapshot, or image, of data stored on a mass storage system (104). At the start of the method, a preservation memory (106) is cleared and a virtual device is created. Whenever a write is to be performed on the mass storage system (104), a check is made of the preservation memory (106) to determine if it contains a block associated with the mass storage write address. If there is not, a copy of the block in the mass storage system (104) at the block write address is placed in the preservation memory (106). Whenever a read is to be performed on the virtual device, a check is made of the preservation memory (106) to determine if it contains a block associated with the virtual device read address. If there is such a block, that block is returned as the result of the virtual device read. Otherwise, the block at the virtual device block read address is returned as the result.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

A STATIC SNAPSHOT OF DATA ON MASS STORAGE

Background of the Invention

Field of the Invention. This invention relates mass storage systems for digital computers, and in particular to a method for providing a static snapshot or image of a mass storage system.

Description of Related Art. It is desirable during the operation of a computer system with a mass storage system, such as a magnetic disk, to periodically make a backup copy of the data stored on the mass storage system to allow for recovery in the event of a failure of the mass storage system. This is commonly done by reading the data stored on the mass storage system and writing it to a magnetic tape.

However, if the data stored on the mass storage system is being updated by other programs as the backup copy is being made, the image of the data on the mass storage system written to tape may be inconsistent. This is because normal backup techniques either copy the blocks from the mass storage system sequentially to the linear-access tape, or walk the file system stored on the mass storage system, starting with the first block of the first file in the first directory and proceeding in order to the last block of the last file of the last directory. The backup program is not aware of updates performed to a block of the mass storage system after that block has been written to tape.

This problem of inconsistent data being written to tape is particularly likely to occur if the mass storage system is being used by a database management system, where an update may involve changing information stored on different parts of the mass storage system. If a database update is made while the backup tape is being written, the image of the database management system written to tape

will have the old values for any data already written to tape at the time of the database update, and the new values for any data written to tape following the database update. A restoration based on the tape image of the database would yield an inconsistent database.

Horton et al., United States Patent No. 5,089,958, which is hereby incorporated by reference in its entirety for the material disclosed therein, discloses a technique for producing an image of a mass storage system at any point in time after the technique is started. This is done by establishing a base image of the mass storage system at the start of the technique and a log indicating each change made to the mass storage system. An image at any point in time can then be produced by starting with the base image and making all the changes indicated in the log up to that point in time. To improve performance, the Horton system also provides for differential images so that the compilation of changes to form an image does not have to start with the base image.

There are two difficulties with using the technique of Horton to provide an image for backup operations. First, the technique is not designed to provide a static snapshot or image of the mass storage system, but to allow an image from any point in time to be created at some later time. This increases the complexity of the technique and requires the compilation of changes whenever a virtual image is desired.

The second difficulty with using the technique of Horton is that the log must store a copy of each change made to the mass storage system in order to produce an image of the mass storage system as it was at a specified time. This means that the size of the log can grow without bound, eventually exhausting the space available for its storage. At this point, updates to the mass storage system

are not possible without compromising the ability to produce an image from any previous point in time.

With many database systems or file systems, certain key blocks (such as master directory blocks) are frequently updated, perhaps with every update to any other block. A copy of these blocks must be written to the log each time they are changed. This will, of course, result in a very large log file, with many of the entries being copies of the key blocks as they changed over time.

Another approach to creating a static image of a mass storage system is possible if the mass storage system has the ability to produce a mirror, or identical copy, of one disk's data on a second disk. At the time the static image is needed, mirroring of data is stopped and the mirror disk is used as the static image. When the static image is no longer necessary (for example, when the tape backup has been completed), the two disks are resynchronized, by copying any changes made during the time mirroring was not active to the mirror disk, and mirroring is resumed.

This approach also has problems. Unless there are three or more disks mirroring the information on the main disk, when mirroring is stopped to produce the static image there is no longer the redundancy of mirrored disk or disks and updates can be lost if there is a disk failure. Furthermore, it requires an entire disk to be devoted to the storage of the static image.

But the major disadvantage of this mirror disk approach is the time necessary to restart mirroring after the static image is no longer needed. This requires updating the mirror disk with all the changes that have been made since mirroring was stopped. If a log of these changes is not available, this means that all the data on the mirror disk must be copied from the disk which has been

updated. For large disks such as would be found on a database system, this could take many hours.

For more general background reference materials, the reader is directed to United States Patent Nos.

3,444,528, 3,533,082, 4,141,066, 4,156,901, 4,164,017,
4,191,996, 4,351,023, 4,378,588, 4,403,303, 4,453,215,
4,459,658, 4,479,214, 4,483,001, 4,484,275, 4,486,826,
4,498,145, 4,507,751, 4,521,847, 4,581,701, 4,607,365,
4,608,688, 4,639,856, 4,648,031, 4,652,940, 4,654,819,
4,654,857, 4,674,038, 4,703,421, 4,703,481, 4,713,811,
4,727,516, 4,736,339, 4,750,177, 4,754,397, 4,878,167,
5,307,481, 5,235,700, 5,079,740, 4,823,256, 5,295,258,
5,157,663, 4,471,429, 4,530,052, 4,615,001, 4,941,087,
4,959,768, 4,979,108, 4,800,488, 4,656,596, 4,866,707,
3,771,137, 4,402,046, 4,423,414, 4,430,699, 4,477,882,
4,480,304, 4,562,535, 4,604,690, 4,630,224, 4,644,470,
4,648,061, 3,754,211, 3,824,547, 4,439,859, 4,445,214,
4,691,314, 3,754,211, 4,332,027, 4,516,121, 4,583,089,
4,646,300, 5,060,185, 5,123,099, 3,602,900, 3,665,173,
3,681,578, 3,735,356, 3,760,364, 3,761,884, 3,810,119,
3,828,321, 3,054,560, 3,303,474, 3,544,477, 3,623,014,
3,636,331, 3,725,861, 3,803,568, 4,012,717, 4,076,961,
4,099,241, 4,118,772, 4,358,823, 4,359,718, 4,455,645,
4,477,895, 4,589,090, 4,590,554, 4,610,013, 4,623,883,
4,639,852, 4,654,846, 4,680,581, 3,557,315, 3,636,331,
3,810,121, 3,812,469, 3,820,085, 3,828,321, 3,864,670,
3,865,999, 3,889,237, 3,959,638, 3,991,407, 4,073,005,
4,099,241, 4,152,764, 4,208,715, 4,228,503, 4,257,009,
4,270,167, 4,282,572, 4,318,173, 4,358,823, 4,371,754,
4,403,286, 4,428,044, 4,455,601, 4,530,051, 4,590,554, and
4,628,508 each of which is hereby incorporated by reference
in their entirety for the material disclosed therein. The
reader's attention is also directed to the following
publications: Lyon, "Tandem's Remote Data Facility," IEEE

(1990); and Molina et al., "Issues in Disaster Recovery," IEEE (1990), each of which is incorporated by reference in its entirety for the material disclosed therein.

Summary of the Invention

It is an object of this invention to provide a static image of data stored on a mass storage system as it existed at a particular point in time.

This is accomplished by creating a virtual device that will appear as a mass storage device containing the static image. Write operations to the mass storage system are also intercepted by the method. Copies of blocks on the mass storage system are placed in a preservation memory whenever they are going to be changed by a write operation, unless an entry for that block is already in the preservation memory. During a read of the virtual device, the preservation memory is first checked, either directly or using a table of contents of the preservation memory, to see if it contains a copy of the block from the specified location. If the preservation memory has such a copy, that copy is returned as the result of the read. Otherwise, the block is read from the mass storage system.

It is a further object of the invention to reduce the amount of storage required to provide the static image. The technique of Horton requires the storage of all changes from the time the technique is started. The mirror disk technique requires storage equal to the size of the mass storage being imaged. In contrast, the method of the invention only requires storage equal to the number of mass storage blocks that have been changed since the static image was created.

It is a further object of the invention to reduce the time necessary for generating the static image and for returning to normal operation when the static image is no longer needed. Unlike the technique of Horton, where the

static image at a particular time needs to be compiled from the base image and log entries, all that is necessary for creating a static image using the method of this invention is to create the virtual device and establish the interception of writes to the mass storage system. No copying of data or compilation of an image is necessary.

When the static image is no longer necessary, the virtual device is removed from the system and the contents of the preservation memory deleted if recovery of that space is desirable. No synchronization to the mass storage system is necessary nor is it necessary to merge updates into a base image to create a new base image.

These and other features of the invention will be more readily understood upon consideration of the attached drawings and of the following detailed description of those drawings and the presently preferred embodiments of the invention.

Brief Description of the Drawings

Figure 1 is a block diagram of a representative digital computer configuration on which the preferred embodiment of the invention operates.

Figure 2 is a flow diagram showing the preferred steps of the method of the invention.

Figure 3 is a block diagram of a client-server configuration using the preferred method.

Figure 4 is a block diagram of the currently-preferred embodiment of the method in a client-server configuration.

Figure 5 is variant of the configuration of Figure 4.

Detailed Description of the Invention

Referring to Figure 1, which illustrates a representative computer configuration on which the method of the invention runs, it can be seen that digital computer 102 has connected to it mass storage system 104 and preservation memory 106. In some embodiments of the invention, association memory 108 may also be connected to digital computer 102.

Mass storage system 104 can be any writable block-addressable storage system, such as one or more disks or a partition of a disk. (If mass storage system 104 were not writable, such as a CD-ROM, its contents would not change and there would be no need for the invention of this application.) A partition of a disk can be a fixed area of a disk. The disks can store their information using magnetic, optical, or any other technique that allows writing and reading of data without departing from the scope and spirit of this invention.

In the currently-preferred embodiment of the invention, digital computer 102 is a PC-compatible computer based on an Intel X86 series or compatible processor and mass storage device 104 is a SCSI or IDE magnetic disk connected to digital computer 102 through an appropriate controller.

Preservation memory 106 can be an area in the random-access memory (RAM) of digital computer 102, one or more disks, a partition of a disk, or a file stored on a disk. Optimal selection of the implementation of preservation memory 106 depends of the number of blocks of the mass storage system that will be changed during the time the static image is needed. The use of RAM provides faster performance, but may limit the number of entries in the preservation memory. If the preservation memory runs out of space when a new entry must be stored, the method fails and the static image is no longer available or

remains in the state it was at the time the preservation memory ran out of space. It is important to note that if this occurs, no data from the mass storage system is lost, and the method can be restarted to produce a new static image.

Referring to Figure 2, which is a flow diagram showing the steps of the method, the method starts at step 202 when a static image of the mass storage system is desired. This can be indicated by the running of a special program, an operating system call, or an operator command, as appropriate for the operating system and application.

In step 202, preservation memory 106 is cleared. In general, this will consist of setting the control information describing the contents of preservation memory 106 to indicate that there are no valid entries in preservation memory 106.

In step 204, a virtual device appearing as a mass storage device is created. The method for creating a virtual device will depend on the particular operating system running on digital computer 102, but will be known by one skilled in the art of that particular operating system. In addition, it may be necessary in step 204 to configure the operating system so that the method of this invention intercepts any read or write operation directed to mass storage system 104. Again, how this is done will be dependent on the particular operating system running on digital computer 102.

In step 206, the method waits until there is a write operation directed to mass storage system 104 or a read operation directed to the virtual device created in step 204. In variants of the method, step 206 also reacts to a read operation directed to mass storage system 104 or a write operation directed to the virtual device created in step 204.

If the operation is a write to mass storage system 104, step 210 is entered. Using the mass storage write address specified in the write operation, step 210 determines if there is a block of data associated with that mass storage write address in preservation memory 106. If there isn't, step 212 of the method is executed. Otherwise, step 212 is skipped and step 214 is executed.

There are a number of ways for determining whether there is a block of data associated with the mass storage address in the preservation memory 106. In the currently-preferred embodiment of the invention, there is a block association memory 108 also connected to digital computer 102. (Block association memory 108 may be a separate memory connected to digital computer 102, or may be a portion of the RAM of digital computer 102.) Block association memory 108 is used to associate blocks stored in preservation memory 106 with the unique addresses of blocks on mass storage system 104. Block association memory 108 does this by containing entries indicating a unique address and the location in preservation memory 106 for the block associated with that unique address.

Entries in block association memory 108 can be stored unordered, in which case they must be linearly searched for a matching unique address. If no entry is found with a matching address, there is not a block in preservation memory 106 associated with that address. Alternatively, the entries could be stored ordered by unique addresses, in which case a binary search could be used to locate the matching entry in block association memory 108. A hashing scheme could also be used to find a matching entry.

The block association memory 108 can also be organized as an array with an element for each unique address of mass storage system 104. Each element of the

array stores a preservation memory location, or a special value that indicates that there is not a block in preservation memory 106 associated with that unique address.

The selection of a technique for storing entries in block association memory 108 depends on the characteristics of accessing the entries. Using an array provides the highest speed for accessing an entry or adding an entry corresponding to a block just copied into preservation memory 106, at the expense of a large block association memory 108. Ordering the entries by unique address provides faster access than for unordered entries, but requires more time when an entry is added to block association memory 108. In the currently-preferred embodiment, entries are stored unordered in block association memory 108.

In this discussion, the term block refers to the data stored at a particular location in mass storage system 104 or preservation memory 106. Blocks are generally of a fixed size (e.g. 512 bytes for disks used with MS-DOS), although blocks of different sizes, or variable sizes, are within the scope of this invention. On mass storage system 104, each block has a unique address, specified in read or write operations. A block in preservation memory 106 is a copy of a block of data stored in mass storage system 104, and that block in preservation memory 106 is associated with the unique address of the block in mass storage system 104 of which it is a copy.

For efficiency, it may be convenient to treat one or more contiguous blocks on mass storage system 104 as if it were a single, large block. Often operating systems perform their mass storage operations on contiguous blocks (called clusters in MS-DOS). The extensions to handle

clusters of blocks should be clear to one with ordinary skills in computer programming.

If block association memory 108 is being used, step 202 (clearing preservation memory 106) consists of removing all entries from block association memory 108 or setting them to the special entry that indicates that there is no block in preservation memory 106 associated with each unique address.

Returning to Figure 2, step 212 is executed if there is not a block associated with the mass storage write address in preservation memory 106. Step 212 places a copy of the block of data currently located at the mass storage write address in preservation memory 106, updating block association memory 108 as necessary. It is important to note that step 212 will be executed at most once for each unique address on mass storage system 104, since the next time step 210 tests to see if there is a block in preservation memory 106 associated with that mass storage write address it will find the copy made by step 212. Because of this, preservation memory 106 will contain only copies of blocks as they were when the method was started.

In step 214, the data to be written by the mass storage write operation is written to the location on mass storage system 104 specified by the mass storage write address. This completes the steps for a mass storage write, and step 206 is reentered to wait for the next operation.

If the operation is a virtual device read, step 220 is entered. Again, a check is made to determine if a block associated with the virtual device read address is in preservation memory 106. If there is such a block, step 224 is executed. If not, step 222 is executed.

Step 222 returns the data from the block in mass storage system 104 specified by the virtual device read

address as the result of the read operation. Step 224 returns the block from preservation memory 106 associated with the virtual address read address as the result of the read operation. This completes the steps for a virtual device read, and step 206 is reentered to wait for the next operation.

If the operation is a mass storage read, step 230 is entered, which returns the data from the block of mass storage system 104 specified by the mass storage read address as the result of the read operation. This completes the steps for a mass storage read, and step 206 is reentered to wait for the next operation.

It may be desirable to allow write operations to the virtual device, changing the image as specified by the write operations. For example, it may be necessary to write a different label or other control information on the virtual device image so the operating system can differentiate it from mass storage system 104.

If the operation is a virtual device write, step 240 is entered. Step 240 checks to see if the virtual device is read-only, and if it is step 242 is entered to return an appropriate error indication to the operating system or user.

Step 244 checks to determine if a block associated with the virtual device write address is in preservation memory 106. If there is such a block, step 248 is executed. If not, step 246 is executed. In step 246, the data from the virtual device write operation is placed in preservation memory 106, associated with the virtual device write address from the virtual device write operation. Block association memory 108 is updated as necessary. In step 248, the data from the virtual device write operation replaces the block associated with the virtual device write address of the virtual device write

operation. This completes the steps of the virtual device write, and step 206 is reentered to wait for the next operation.

While the description above describes the basic operation of the method of the invention, there are a number of other embodiments possible. For example, the same preservation memory 106 can be shared so that a second virtual device provides a snapshot image of a second mass storage system. In another embodiment, a second preservation memory and second virtual device can be used to provide a second image whose snapshot was taken at a different time of mass storage system 104.

The computer system running the method of the invention can also be used as a file server for client computers connected to it by a network or other means. As a file server, it can export its mass storage system, the virtual device created by the method, or both. Such a system is illustrated in Figure 3.

File server computer 312, with mass storage system 314, runs the method of the invention. It exports the virtual device (and probably mass storage system 314) to client computer 302, communicating over network 300. Computer 302 can run a tape backup program that copies the information from the exported virtual device to tape drive 304. No change is necessary for the tape backup program running on client computer 302, which sees the virtual device as just another mass storage device.

Figure 4 illustrates the currently-preferred configuration for running the method of the invention. Network 400 connects client computer 402, with tape drive 404, to file server computer 412, with mass storage system 414. File server computer communicates with standby server computer 422 over data link 420. Standby server computer 422 has mass storage system 424. Through software running

on file server computer 412 and standby server computer 422, as described in United States Patent application serial number 08/094,744, filed on July 20, 1993 and entitled "METHOD FOR RAPID RECOVERY FROM A NETWORK FILE SERVER FAILURE" (which is hereby incorporated by reference in its entirety), mass storage system 424 appears as a disk to file server computer 412 and mirrors the data on mass storage system 414. In the event of a failure of either file server computer 412 or mass storage system 414, standby computer 422 can be restarted as the file server.

In the configuration of Figure 4, standby server 422 runs the method of the invention, and can export the virtual device either to file server computer 412, which can then export it to client computers on network 400, or standby server 422 can directly export the virtual device to client computers. The virtual device can also be accessed by programs running on standby server 422.

Figure 5 illustrates a variant of the configuration of Figure 4. Instead of client computer 502 having a tape drive, as was the case for client computer 402, backup computer 532 has tape drive 534. Backup computer 532 communicates with standby server computer 522 over data link 530. Standby server computer exports mass storage system 524 to file server computer 512 (whether mirrored or not). Standby server computer exports the virtual device with the snapshot image of mass storage system 524 to backup computer 532.

Backup computer 532 can now copy the snapshot image of mass storage system 524 by reading the virtual device exported to it by standby server computer 522. Neither file server computer 512 nor standby server computer 522 has the overhead of the tape backup process, which can result in a degradation of performance if data compression needs to be performed before the data is

written to tape. Also, a fault in the tape backup program will not affect either file server computer 512 or standby server computer 522.

It is to be understood that the above described embodiments are merely illustrative of numerous and varied other embodiments which may constitute applications of the principles of the invention. Such other embodiments may be readily devised by those skilled in the art without departing from the spirit or scope of this invention and it is our intent they be deemed within the scope of our invention.

Claims

We claim:

1. A method for providing a static snapshot of data stored on a mass storage system, operating on a computer configuration including:

- a digital computer executing the steps of the method;
- a mass storage system connected to said digital computer, said mass storage system storing blocks of data having unique addresses; and
- a preservation memory connected to said digital computer, said preservation memory storing blocks of data associated with said unique addresses;

the method comprising:

- (A) clearing said preservation memory so that no copies of blocks of data are in said preservation memory;
- (B) creating a virtual device;
- (C) whenever a write operation to said mass storage system, said write operation specifying a mass storage write address and data to be written, occurs:
 - (1) if there is not a block of data associated with said mass storage write address in said preservation memory, placing a copy of said block of data located in said mass storage system at said mass storage write address in said preservation memory; and
 - (2) writing said data to be written to said mass storage system at the location specified by said mass storage write address;

and

(D) whenever a read operation to said virtual device, said read operation specifying a virtual device read address, occurs:

- (1) if there is not a block of data associated with said virtual device read address in said preservation memory, returning said data of said block of said mass storage system specified by said virtual device read address as the result of said read operation, and
- (2) if there is a block of data associated with said virtual device read address in said preservation memory, returning said block of data from said preservation memory as the result of said read operation.

2. A method as in claim 1, wherein said mass storage system comprises one or more disks.

3. A method as in claim 1, wherein said mass storage system is a partition of a disk.

4. A method as in claim 1, wherein said preservation memory is a random-access memory.

5. A method as in claim 1, wherein said preservation memory is one or more disks.

6. A method as in claim 1, wherein said preservation memory is a partition of a disk.

7. A method as in claim 1, wherein said preservation memory is a file stored on mass storage system.

8. A method as in claim 1, the method further comprising:

- (E) whenever a write operation to said virtual device, said write operation specifying a

virtual device write address and data to be written, occurs:

- (1) if there is not a block of data associated with said virtual device write address in said preservation memory, placing in said preservation memory said data to be written, and
- (2) if there is a block of data associated with said virtual device write address in said preservation memory, replacing in said preservation memory that block of data with said data to be written.

9. A method as in claim 1, the method further comprising:

- (F) whenever a read operation to said mass storage system, said read operation specifying a mass storage read address, occurs, returning said data of said block of said mass storage system specified by said mass storage read address as the result of said read operation.

10. A method as in claim 1, said computer configuration further including a block association memory, said block association memory used to associate blocks stored in said preservation memory with said unique addresses.

11. A method as in claim 10, wherein said block association memory contains entries indicating a unique address in said mass storage system and a location in said preservation memory of a block associated with that unique address.

12. A method as in claim 11, wherein blocks of data in said preservation memory are associated with a unique

address by searching said block association memory entries for a matching address.

13. A method as in claim 12, wherein there is not a block of data associated with an address if there is no entry in said block association memory with a matching address.

14. A method as in claim 11, wherein said block association memory contains an entry for each unique address in said mass storage system indicating a location in said preservation memory of a block associated with that unique address.

15. A method as in claim 12, where a special value for said preservation memory location in said entries indicates that there is not a block of data in said preservation memory associated with that address.

16. A method as in claim 1, wherein said digital computer acts as a file server, and said virtual device is exported to other computers.

17. A method as in claim 16, wherein said mass storage system is exported to other computers.

18. A method as in claim 1, said computer configuration including a second mass storage system, the method further comprising:

creating a second virtual device;

whenever a write operation to said second mass storage system, said write operation specifying a second mass storage write address and data to be written, occurs:

- (1) if there is not a block of data associated with said second mass storage write address in said preservation memory, placing a copy of said block of data located in said second mass storage system at said

second mass storage write address in said preservation memory; and

- (2) writing said data to be written to said second mass storage system at the location specified by said second mass storage write address;

and

whenever a read operation to said second virtual device, said read operation specifying a virtual device read address, occurs:

- (1) if there is not a block of data associated with said virtual device read address in said preservation memory, returning said data of said block of said second mass storage system specified by said virtual device read address as the result of said read operation, and
- (2) if there is a block of data associated with said block read address in said preservation memory, returning said block of data from said preservation memory as the result of said read operation.

19. A method as in claim 1, said computer configuration including a second preservation memory, the method further comprising:

creating a second virtual device;

whenever a write operation to said mass storage system, said write operation specifying a mass storage write address and data to be written, occurs:

- (1) if there is not a block of data associated with said mass storage write

address in said second preservation memory, placing a copy of said block of data located in said mass storage system at said mass storage write address in said second preservation memory; and

- (2) writing said data to be written to said mass storage system at the location specified by said mass storage write address;

and

whenever a read operation to said second virtual device, said read operation specifying a second virtual device read address, occurs:

- (1) if there is not a block of data associated with said second virtual device read address in said second preservation memory, returning said data of said block of said mass storage system specified by said second virtual device read address as the result of said read operation, and
- (2) if there is a block of data associated with said second virtual device read address in said second preservation memory, returning said block of data from said second preservation memory as the result of said read operation.

20. A method for providing a static snapshot of data stored on a mass storage system, the method operating on a computer configuration that includes:

- a digital computer;
- a mass storage system connected to said digital computer, said mass storage system being

- capable of storing blocks of data having unique addresses; and
- a preservation memory connected to said digital computer, said preservation memory being capable of storing blocks of data associated with said unique addresses;
- the method comprising the following steps:
- (A) clearing said preservation memory so that no copies of blocks of data are in said preservation memory;
 - (B) creating a virtual device;
 - (C) whenever a write operation to said mass storage system occurs that specifies a mass storage write address and data to be written occurs, performing the following:
 - (1) if there is not a block of data associated with said mass storage write address in said preservation memory, placing a copy of said block of data located in said mass storage system at said mass storage write address in said preservation memory; and
 - (2) writing said data to be written to said mass storage system at the location specified by said mass storage write address;
- and
- (D) whenever a read operation to said virtual device that specifies a virtual device read address occurs, performing the following:
 - (1) if there is not a block of data associated with said virtual device read address in said preservation memory, returning said data of said

block of said mass storage system specified by said virtual device read address as the result of said read operation, and

- (2) if there is a block of data associated with said virtual device read address in said preservation memory, returning said block of data from said preservation memory as the result of said read operation.

21. A method for providing a static snapshot of data stored on a mass storage system, the method operating on a computer configuration that includes:

- a digital computer;
- a mass storage system connected to said digital computer, said mass storage system being capable of storing blocks of data having unique addresses; and
- a preservation memory connected to said digital computer, said preservation memory being capable of storing blocks of data associated with said unique addresses;

the method comprising the following steps:

- (A) clearing said preservation memory so that no copies of blocks of data are in said preservation memory;
- (B) creating a virtual device;
- (C) whenever a write operation to said mass storage system occurs that specifies a mass storage write address and data to be written occurs, performing the following:
 - (1) if there is not a block of data associated with said mass storage write address in said preservation memory,

- placing a copy of said block of data located in said mass storage system at said mass storage write address in said preservation memory; and
- (2) writing said data to be written to said mass storage system at the location specified by said mass storage write address;
- (D) whenever a read operation to said virtual device that specifies a virtual device read address occurs, performing the following:
- (1) if there is not a block of data associated with said virtual device read address in said preservation memory, returning said data of said block of said mass storage system specified by said virtual device read address as the result of said read operation, and
- (2) if there is a block of data associated with said virtual device read address in said preservation memory, returning said block of data from said preservation memory as the result of said read operation;
- (E) whenever a write operation to said virtual device occurs that specifies a virtual device write address and data to be written, performing the following:
- (1) if there is not a block of data associated with said virtual device write address in said preservation memory, placing in said preservation memory said data to be written, and

- (2) if there is a block of data associated with said virtual device write address in said preservation memory, replacing in said preservation memory that block of data with said data to be written; and

- (F) whenever a read operation to said mass storage system that specifies a mass storage read address occurs, returning said data of said block of said mass storage system specified by said mass storage read address as the result of said read operation.

22. A system for providing a static snapshot of data stored on a mass storage system on a computer configuration that includes:

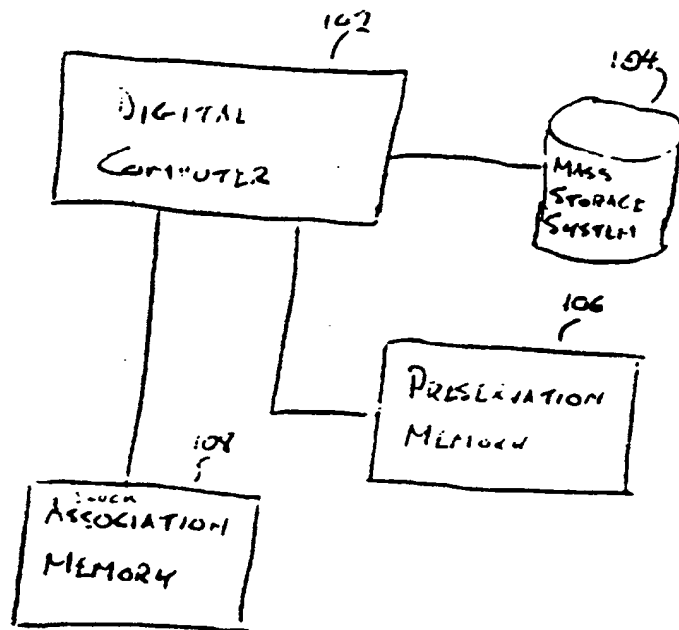
- a digital computer;
- a mass storage system connected to said digital computer, said mass storage system being capable of storing blocks of data having unique addresses; and
- a preservation memory connected to said digital computer, said preservation memory being capable of storing blocks of data associated with said unique addresses;

the system comprising:

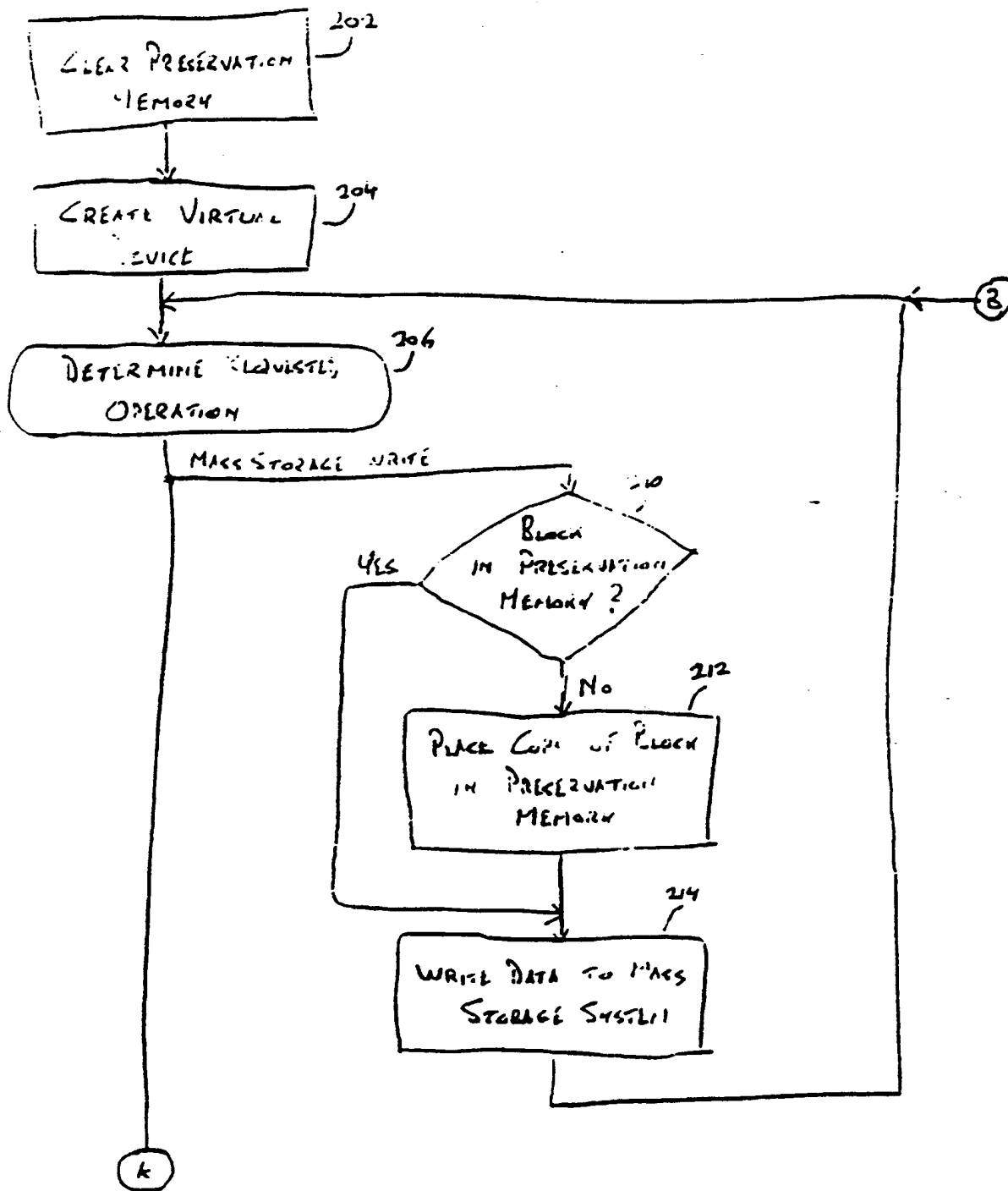
- (A) means for clearing said preservation memory so that no copies of blocks of data are in said preservation memory;
- (B) means for creating a virtual device;
- (C) whenever a write operation to said mass storage system occurs that specifies a mass storage write address and data to be written occurs, means for performing the following:

- (1) if there is not a block of data associated with said mass storage write address in said preservation memory, placing a copy of said block of data located in said mass storage system at said mass storage write address in said preservation memory; and
 - (2) writing said data to be written to said mass storage system at the location specified by said mass storage write address;
- (D) whenever a read operation to said virtual device that specifies a virtual device read address occurs, means for performing the following:
 - (1) if there is not a block of data associated with said virtual device read address in said preservation memory, returning said data of said block of said mass storage system specified by said virtual device read address as the result of said read operation, and
 - (2) if there is a block of data associated with said virtual device read address in said preservation memory, returning said block of data from said preservation memory as the result of said read operation;
- (E) whenever a write operation to said virtual device occurs that specifies a virtual device write address and data to be written, means for performing the following:

- (1) if there is not a block of data associated with said virtual device write address in said preservation memory, placing in said preservation memory said data to be written, and
 - (2) if there is a block of data associated with said virtual device write address in said preservation memory, replacing in said preservation memory that block of data with said data to be written;
and
- (F) whenever a read operation to said mass storage system that specifies a mass storage read address occurs, means for returning said data of said block of said mass storage system specified by said mass storage read address as the result of said read operation.

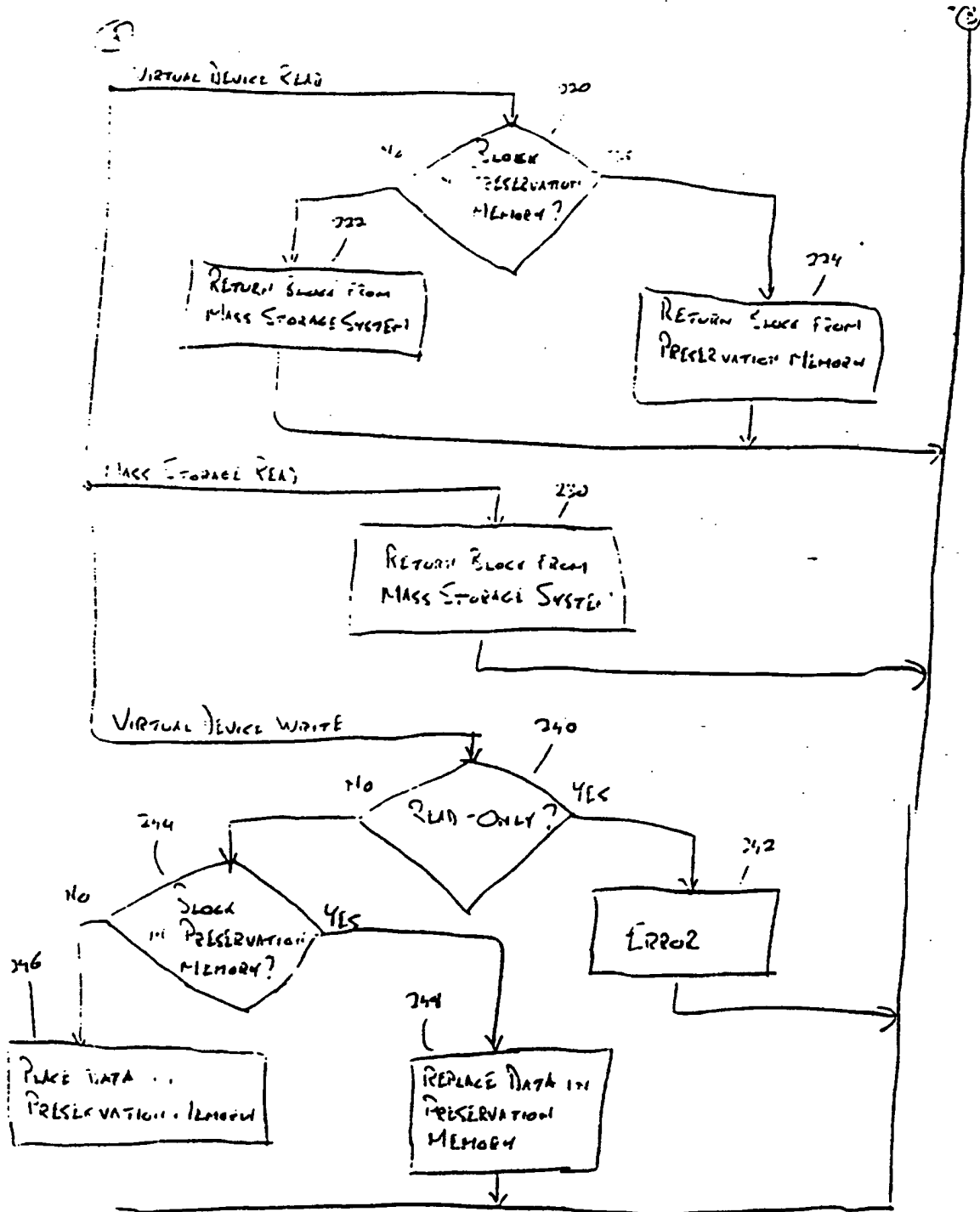


- FIGURE 1 -



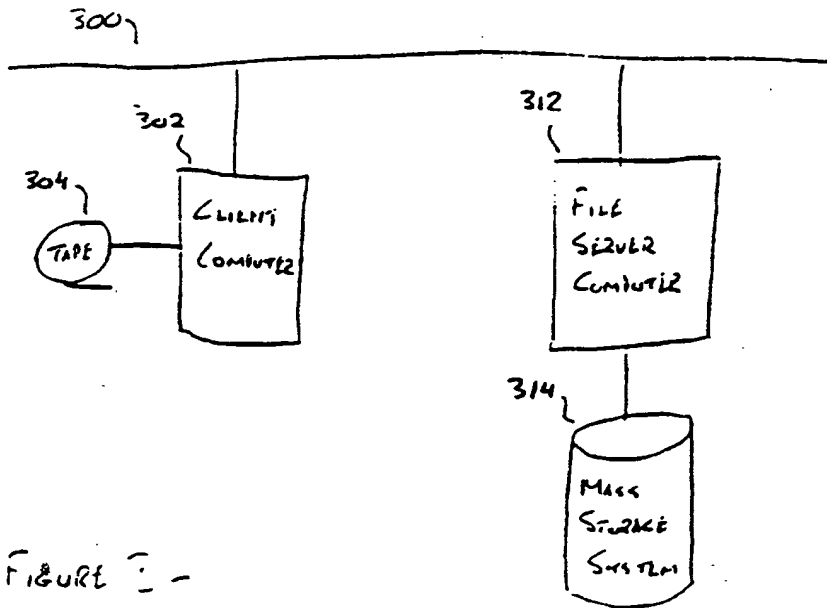
- FIGURE 2 -

3/5

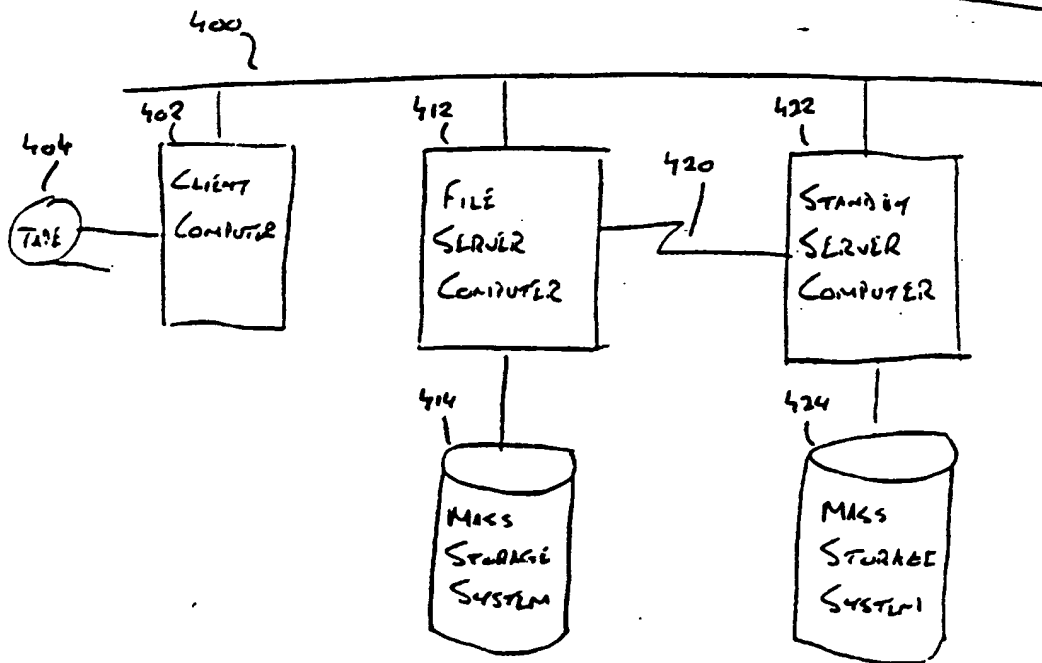


- FIGURE 2 [CONT] -

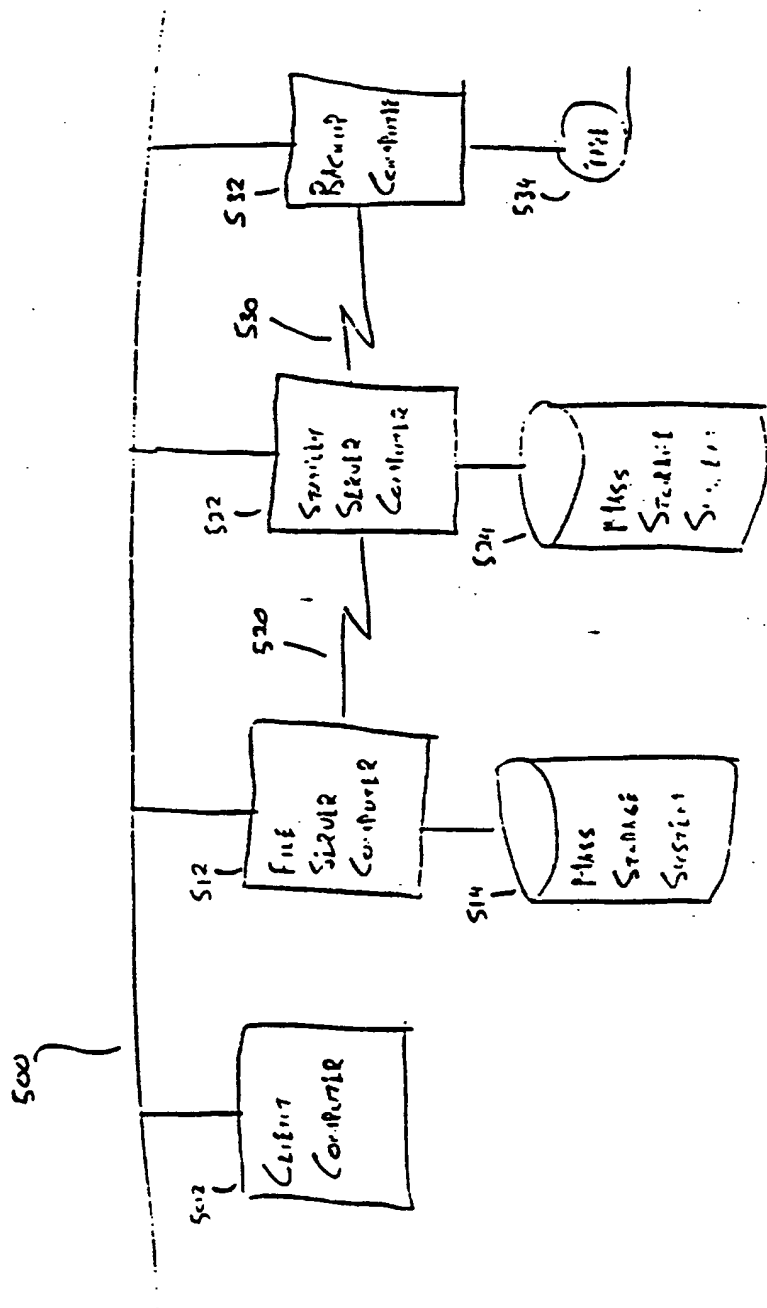
4/5



- FIGURE 3 -



- FIGURE 4 -



-figure 5-

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/13324**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) : G06F 12/16, 12/08

US CL : 395/468, 488, 462, 412

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 395/468-470, 488-489, 462, 412-413, 417 and 419.

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	US, A, 5,379,391 (BELSAN ET AL) 03 January 1995, column 1, lines 17-22, 43-49, column 11, line 1 to column 13, line 39. Also, see the abstract.	1-14, 18-22
X,P	US, A, 5,426,747 (WEINREB ET AL) 20 June 1995, column 2, lines 50-60, column 6, line 57 to column 7, line 29, claim 1, column 11, lines 30-38, column 12, lines 10-15, column 14, lines 29-37, column 3, lines 17-21.	1-17, 20-22



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principles or theory underlying the invention
A document defining the general state of the art which is not considered to be part of particular relevance	X	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
E earlier document published on or after the international filing date	Y	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	A	document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means		
P document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

05 JANUARY 1996

Date of mailing of the international search report

22 MAR 1996

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer
ARIO ETIENNE

Telephone No. (703) 308-7562

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/13324

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

USPTO, Automated Patent Search (APS),
Files: USPAT & JPOAPS

Search terms: snapshot, image, intercept, virtual device, disks, mass storage, cache, flush, clear, files.